THE UNIVERSITY
of EDINBURGH

Functional Characterization of the DNA Glycosylase; Methyl-CpG Binding Domain Protein 4 (MBD4)

Meng Huan

Thesis presented to the University of Edinburgh
for the Degree of Doctor of Philosophy
2012

Declaration

I declare that this thesis has been composed by me, and that all of the work is my own unless otherwise stated.

Meng Huan
May 2012

Acknowledgements

First of all I'd like to thank Richard (Professor Richard R. Meehan) for giving me the chance to do my PhD project in his lab and for his guidance, encouragement, and enthusiasm throughout three and half years. Thank you also to David (Professor David J. Harrison) who is also my PhD supervisor has provided advice and support throughout this project in addition to his help for the funding of my extension period. Thank you to my graduate studies panel members, Bob Hill and Colin Gordon, for their time, advice and opinions during my progress meetings. Thank you also to the medical college of Edinburgh University for financial support, in particular, the initial recommendation and support from David (Professor David J. Porteous) for my PhD study at Edinburgh.

Thank you to my examiners Doctor Bernard Ramsahoye and Professor Peter Meyer. It was my pleasure to have my viva with their examination. I learned a lot from their independent opinions and their suggestions and questions. Actually after I listened to and learned from senior scientists like my supervisors, the present and the past, my examiners, my interviewers, and my colleagues, I made sure that I need to look into a specific model/process and/or specific pathways, in which I can develop my view/vision, as well as refine my techniques in my early scientific career.

Thank you to everyone in the Richard' lab, present and past members for their help. In particular thank you to Monika Madej for teaching me the molecular cloning and band shift techniques at very beginning of my project. Thank you to James Reddington for discussions in the lab and his comments of this thesis. Thank you to Inhwa Um in David's lab for helping me the tissue microarray work. Thank you to John, Diana, Colm, Raffy, Sara and Michael for discussions and making the lab an enjoyable environment.

Thank you to everyone else in the MRC human genetics building for their help, advice, materials and/or friendship. In particular, I'd like to thank Heidi Sutherland and Pradeepa Madapura-Marulasiddappa for their help in Y2H and AP/MS. Shelagh Boyle for help and materials; Stephen Brown, Sean O'Neil, Paul Perry, and everyone else from core scientific services for my huge sequencing, plates, and imaging work. The same goes for my office mates, in particular Bret Heale for scientific discussions, and Xianghua Li, the dear aunt of my daughters. Thank you to friends both in and outside the research hospital for dinners, parties, and interesting conversations regarding culture, history, and philosophy.

Thank you to my parents for their love and support throughout my life. The family faith they pass to me as a Mencius's descendants make me truly believe the power of

education, fulfil my mission for helping others, and always hold compassion for the suffering people. Thank you to my wife Jing for her hard work to take care of our family, our daughters Shuyan and Wenxi. I feel it were just at yesterday I hold your hand walking through our campus 12 years before. THANK YOU!

Thank you to my home China, lovely country and lovely people. They are my root and the source of my energy. I believe that a competitive and stable China will not only benefit the world as a whole, but offer the modern society an opportunity of reappraisal in terms of politics, economy, culture and social system.

Abstract

DNA methylation is a major form of epigenetic modification and involves the addition of a methyl group covalently to the 5-position of the cytosine pyrimidine ring, mostly within the context of CpG dinucleotides in vertebrate somatic cells. Methylation of CpG dinucleotides at promoter regions is generally associated with transcriptional repression. In this context, the methyl-CpG binding proteins (MeCPs) that are capable of recognition of methylated CpG dinucleotides are proposed to play a central role in DNA methylation associated transcriptional repression. Methyl-CpG binding domain protein 4 (MBD4) is an MeCP that possesses a glycosylase domain at its C-terminal, which can excise and repair both G:T and G:U mutations derived from DNA deamination at CpG dinucleotides, in addition to its N-terminal MBD binding domain. MBD4 has been associated with a number of pathways including DNA repair, apoptosis, transcriptional repression, and possibly DNA demethylation processes. However, the precise contribution of MBD4 to these processes remains unclear.

To explore the functional repertoire of MBD4 I decided to undertake multiple protein interaction studies to identify potential partner proteins. I performed yeast 2-hybrid screens with an 11.5 day mouse embryonic cDNA library and multiple mass spectrometry of immunoprecipitates of tagged versions of MBD4 that were over-expressed in human cell lines. I detected ~380 potential interacting candidates with these assays. A significant number of candidates were detected in both assay systems. Chosen candidates were further validated by reciprocal co-IP of expressed partners and by immunofluorescence (IF) microscopy to determine their potential co-localisation in mouse and human cell lines. Subsequently, I identified the intervening domain of MBD4 as a novel protein interaction region for tested candidates. My analysis suggests that MBD4 can have a role in regulation of post-replication methyl-error repair/methylation machinery through its direct interaction with DNMT1 (previously shown), UHRF1 (novel) and USP7 (novel), as well as possible cross-talk to histone modification and chromatin remodelling pathways, through partners such as PRMT5 and ACF1. Interestingly the transcription regulatory components KAP1 and CFP1 not only interact with but also dramatically influence the stability of exogenously expressed MBD4 in human cells. In general positive validation by IP and IF demonstrates the robustness of the initial screens, and implies that MBD4 may impact upon several transcriptional and epigenetic networks along with a number of nuclear pathways that include transcriptional repression, DNA repair and RNA processing.

To test for transcriptional aberration in the absence of *Mbd4* function I profiled two independent mouse cell lines that lack MBD4 activity using Illumina MouseWG-6 v2.0 Expression BeadChip arrays. A number of genes were identified that are significantly up- or down- regulated in both *Mbd4*$^{-/-}$ MEFs. This included mis-expression of insulin-like growth factor-binding proteins and two paternally imprinted genes *Dio3* and *H19*. The cohort of genes that were mis-expressed in the *Mbd4*$^{-/-}$ MEFs overlap with genes that responsed to tamoxifen exposure in an ER-positive ZR-75-1 xenograft model. In response to this observation I identified a potential interaction between MBD4 and estrogen receptor α (ERα) by co-IP and IF co-localisation. This suggests that MBD4 might potentiate transcription of estrogen regulated genes via a direct interaction with ERα, supporting a possible link between replication repair remodelling and steroid/thyroid hormone receptor transcriptional regulation. Additionally I performed a pathway analysis by which several developmental genes including *Sox9*, *Klf2* and *Klf4*, were prioritised as possible MBD4 targets. On this basis I propose a role for MBD4 in acquired diseases such as cancers and autoimmune diseases via transcriptional regulation.

I also performed a comparison of MBD4 DNA binding activity with MBD4 homologues from the Medaka fish (*Oryzias latipes*) and the amphibian, *Xenopus laevis*. I could show that DNA binding specificity to a series of methylated and mismatched probes is conserved regardless of the poor sequence conservation of the MBD domain of MBD4 between the species. I conclude that MBD4 is integrated in multiple pathways in the nucleus that includes DNA repair, chromatin remodelling, transcriptional regulation and genome stability.

List of abbreviations

AML	Acute myeloid leukemia
BSA	bovine serum albumin
ddNTP	dideoxynucleoside triphosphate
dGTP	deoxyguanosine triphosphate
PBS	phosphate-buffered saline
pen	penicillin
SAP	shrimp alkaline phosphatase
TBE	tris-EDTA boric acid buffer
TBS(T)	tris buffered saline (plus 0.05% w/v Tween-20)
˚C	degrees Celsius
3-AT	3-amino-1,2,4-triazole
5 or 3′	5 or 3 prime
5-Aza-dC	5-Aza-2'-deoxycytidine
5-FU	5-fluorouracil
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5'-hydroxymethylcytosine
5mC	5'-methylcytosine
5mCpG	Methylated cytosine in the context of a CpG motif
A	adenine or Ampere
A_{260}	absorption at 260 nm
aa	amino acid(s)
Ac	acetate
AD	activation domain
ADH1	alcohol dehydrogenase 1
Amp	ampicillin
AP	Alkaline phosphatase
AP	apurinic/apyrimidinic
AP	Affinity purification
Apc	adenomatous polyposis coli
APS	ammonium persulphate
ATP	adenosine 5’-triphosphate
ATPase	adenosine triphosphatase
Ax	absorbance at x nm
AZA	5'-deoxy-2'-azacytidine
b	base(s)
BAC	Bacterial artificial chromosome
BAH	bromo adjacent homology
BER	base excision repair
BLAST	basic local alignment search tool
BMP	bone morphogenetic protein
bp	base pair(s)
BPB	bromophenol blue
BS	Bisulfite sequencing
BSA	bovine serum albumin
C	carboxy

C-terminal	Carboxyl terminal (referring to polypeptide)
C, G, A, T	Cytosine, guanosine, adenine and thymine
C. elegans	*Caenorhabtidis elegans*
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CGI	CpG island
ChIP	chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation on a microarray (chip)
cm	centimetre
Cm	chloramphenicol
CpG	cytosine and guanine seperated by a phosphate
cpm	counts per minute
cRNA	complementary ribonucleic acid
CTP	cytidine 5'-triphosphate
CXXC	refers to amino acid motif (cysteine, x, x, cysteine)
Cy	Cyanine
d	day(s)
Da	daltons
DAPI -	4',6-diamidino-2-phenylindole
dATP	deoxyadenosine 5'-triphosphate
DBD	DNA binding domain
dCTP	deoxycytidine 5'-triphosphate
ddATP	dideoxyadenosine triphosphate
ddCMP	dideoxycytidine monophosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
DDO	double dropout
ddTTP	dideoxythymidine triphosphate
DEPC	diethylpyrocarbonate
dGTP	deoxyguanosine 5'-triphosphate
dH20	distilled water
DMEM	dulbecco's modified eagle medium
DMF	dimethylformamide
DMR	differentailly methylated region
DMR	double mismatch reversion
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
Dnmt	dna methylatransferase
Dnmt3	A protein of the DNA methyltransferase 3 family
dNTP	deoxynucleoside 5'-triphosphate
dpp	days post partum
Drosophila	*Drosophila melanogaster*
ds	double-stranded
dsDNA	Double-stranded DNA
DTT	dithiothreitol
dTTP	deoxythymidine 5'-triphosphate
E	embryonic day
E. coli	Escherichia coli
EB	embryoid body
EC	embryonal carcinoma (cell)

ECL	enhanced chemi-luminescence
EDTA	ethylenediaminotetra-acetic acid
EG	embryonic germ (cell)
EGFP	enhanced GFP
EMBL	European Molecular Biology Laboratory
EpiSC	epiblast stem cell
ES	embryonic stem (cell)
EtBr	ethidium bromide
FADD	Fas-associated death domain protein
FBS	foetal bovine serum
FCS	foetal calf serum
FDR	False discovery rate
FGF	fibroblast growth factor
Fig.	figure
FISH	Fluorescence in situ hybridisation
FITC	Fluorescein isothiocyanate
g	gram(s)
g	Gravitational constant
Gal	D-galactose
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
Glu	D-glucose
GluRS	glutamyl tRNA synthetase
GMEM	glasgow's modified eagle medium
GO	Gene ontology
GS	germline stem (cell)
GST	glutathione-S-transferase
GTP	guanosine 5'-triphosphate
h	hour(s)
H	A, C or T (DNA bases)
H3K27	histone H3 lysine 27
H3K27me3	histone H3 (H3), lysine at position 27 (K27), trimethylation (me3)
H3K4	histine H3 lysine 4
H3K9	histone H3 lysine 9
HAT	histone acetyltransferase
HCl	hydrochloric acid
HCP	High-density CpG promoter
HDAC	Histone deacetylase
HELP	HpaII tiny fragment enrichment by ligation-mediated PCR
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HGU	Human genetics unit
HhH	helix-hairpin-helix
His	histidine
HMC	Histone modifying complex
HMT	histone methylatransferase
Hox	Homeotic/homeobox
Hox	homeobox
HP1	heterochromatin associated protein 1
HRP	horse radish peroxidase
IAP	Intercisternal A particle
IC	immunocytochemistry

ICF	Immunodeficiency, centromere instability and facial abnormalities (a syndrome)
ICM	inner cell mass
ICP	Intermediate-density CpG promoter
ICR	Imprinting control region
Ig	immunoglobulin
IgG	Immunoglobulin G
IP	Immunoprecipitation
IPTG	isopropyl-β-D-thiogalactopyranoside
k	kilo
Kan	kanamycin
kb	kilobase
kDa	kilo Dalton
KO or ko	knockout
l	litre(s)
LB	Luria-Bertani
LCP	Low-density CpG promoter
lincRNA	Long non-coding RNA
LINE	Long interspersed nuclear elements
LOH	loss of heterozygosity
M	molar
m	milli (10-3) or metre
M-phase	mitotic phase
mA	milliampere(s)
MBD	Methyl-CpG binding domain
MCS	multiple cloning site
MeCP	methyl-CpG binding protein
MeDIP	Methyl-DNA-immunoprecipitation
MEF	mouse embryonic fibroblast
mES	mouse embryonic stem (cell)
mg	milligram(s)
min	minute(s)
miRNA	micro RNA
ml	millilitre(s)
MLL	Mixed lineage leukemia
mM	millimolar
MMR	mismatch repair
Mnase	Micrococcal nuclease
MNNG	N-methyl-N′-nitro-N-nitrosoguanidine
MOPS	3-[N-morpholino]propanesulphonic acid
mRNA	messenger ribonucleic acid
MSI	microsatellite instability
MW	molecular weight
n	nano (10-9)
N-ChIP	Native chromatin immunoprecipitation
N-CoR	Nuclear co-repressor
N-terminal	Amino terminal (referring to polypeptide)
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
NEM	N-ethylmaleimide
ng	nanogram(s)

NLS	nuclear localisation signal
nm	nanometre(s)
NP-40	nonidet P-40 detergent
nt	nucleotide
NTP	Nucleotide tri-phosphate
o/n	overnight (generally 16-18h)
OD_{600}	optical density at 600 nm
ODxnm	optical density at x nm
Oligo	oligonucleotide
ORF	open reading frame
p	pico (10-12) or plasmid
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBS(T)	phosphate buffered saline (plus 0.2% w/v Tween-20)
PcG	polycomb group proteins
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFA	paraformaldehyde
PGC	primordial germ cell
pH	log10[H+]
PHD	plant homeodomain
PhoRC	Pho-repressive complex
pI	isoelectric point
PIAS	protein inhibitor of activated STAT
piRNA	piwi-interacting RNA
PMD	Partially methylated domain
pmol	picomole(s)
PMSF	phenyl methyl sulfonyl fluoride
PNK	T4 polynucleotide kinase
PolII	RNA polymerase II
PR-DUB	Polycomb repressive de-ubiquitinase
PRC	polycomb repressive complex
PRC1	Polycomb repressor complex 1
PRC2	Polycomb repressor complex 2
pRNA	Promoter RNA
ProK	Proteinase K
PVDF	polyvinylidine difluoride
PWWP	Refers to amino acid motif (proline, tryptophan, tryptophan, proline)
QDO	quadruple dropout
qPCR	Quantitative polymerase chain reaction
qRT-PCR	quantitative reverse-transcriptase PCR
r/t	room temperature
rDNA	Ribosomal DNA
RING	really interesting new gene (protein domain)
RIPA	Radioimmunoprecipitation
RNA	ribonucleic acid
RNase	ribonuclease
rNTP	ribonucleoside 5'-triphosphate
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid

RT	room temperature
RT	Reverse transcriptase
RT-PCR	reverse transcriptase polymerase chain reaction
s	seconds
S-phase	Synthesis (of DNA) phase
S. cerevisiae	*Saccharomyces cerevisiae*
SAM	S-adenosyl-L-methionine
SD-	synthetic defined *or* dropout medium minus
SDS	sodium dodecyl sulphate
SEM	Standard error of the mean
Seq	sequencing (e.g. ChIP-seq)
SINE	Short interspersed nuclear elements
snRNA	small nuclear ribonucleic acid
SRA	SET and RING associated domain
sRNA	small (non-coding) RNA
SSC	sodium choride-sodium citrate (buffer)
SSC	Saline-sodium citrate
ssDNA	single stranded DNA *or* salmon sperm DNA
Strep	streptomycin
SUMO	small ubiquitin-related modifier
SV40	Simian virus 40
SWI/SNF	SWItch/Sucrose NonFermentable
SWIM	SWI2/SNF2 and MuDR
TAE	tris-EDTA acetic acid buffer
TBE	Tris-Borate-EDTA (buffer)
TBS	Tris-buffered saline
TBST	Tris-buffered saline-Tween-20
TDG	T:G mismatch-specific thymine DNA glycosylase
tDMR	Tissue-specific differentially methylated region
TE	Tris-EDTA (buffer)
TE	trophectoderm
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
Tet	tetracycline
TET	Ten-eleven translocation (protein family)
TF	Transcription factor
Tm	melting temperature
TRD	Transcription repression domain
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TSA	trichstatin A
TSS	transcription start site
TxRd	Texas red
U	unit(s)
UAS	upstream activating sequence
UCSC	University of Santa Cruz, California, USA
Ulp1	ubiquitin-like protease 1
Ura	uracyl
USP2cc	Usp2-45 catalytic core
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet

V	volt(s)
v/v	volume per unit volume
vol	volumes
VSP	very short patch
W	Watt(s)
w/o	without
w/v	weight per unit volume
WB	western blot
WT or wt	wild type
X-ChIP	Cross-linked chromatin immunoprecipitation
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Xa	Active X chromosome
XCI	X-chromosome inactivation
Xi	Inactive X chromosome
YMM	yeast minimal medium
YPDA	yeast extract, peptone, glucose (dextrose), and adenine
YPGalA	yeast-peptone-galactose-adenine (medium)
αKG	α-keto gluterate
λ	lambda or wavelength
µ	micro (10-6)
µF	microfarad(s)
µg	microgram(s)
µl	microlitre(s)
µm	micrometre(s)
Ψ	large hydrophobic amino acid

Amino acid abbreviations

Amino acid	Three letter abbreviation	Single letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Iso	I
Leucine	Leu	L

Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

The genetic code

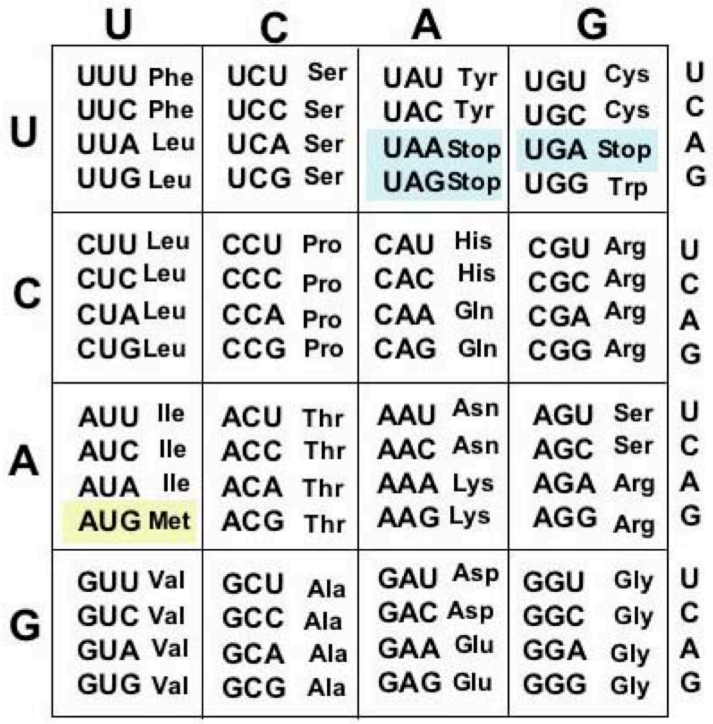


Table of contents

Chapter 1: Introduction

1.1. Overview

One definition of Epigenetics is that it is concerned with alterations in phenotype caused by changes in cellular properties that are inherited but do not represent an alteration in genotype. In vertebrate somatic cells, epigenetic regulation of gene expression is thought to reinforce stable expression states at different loci (Jaenisch and Bird, 2003). These are associated with particular molecular signatures of DNA and chromatin modifications connected with active and repressed chromatin states (Barski et al., 2007). Epigenetic signatures such as DNA methylation are developmentally regulated and are thought to define comparable tissues and differentiation states (Baranzini et al., 2010). In cancer, as well as in embryos generated through somatic cell nuclear transfer, normal patterns of DNA methylation are altered (Sproul et al., 2011, Beaujean et al., 2004). These changes are thought to underlie the molecular pathology of these 'disease' states (Illingworth et al., 2010, Illingworth and Bird, 2009, Beaujean et al., 2004).

This chapter begins by introducing the epigenetic mark of DNA methylation, and the machinery and the biological functions of DNA methylation in vertebrates. The methyl-CpG binding proteins act as interpreter of DNA methylation and the possible link between DNA methylation, repair and the other epigenetic mechanisms that synchronise the transcriptional repression events are then discussed. Finally, the methyl-CpG binding DNA repair glycosylase MBD4 (also known as MED1) is discussed in detail, with a brief introduction of the only other mammalian thymine glycosylase TDG. At the end of the chapter, the aims of the research presented in Chapters 3 to 6 are outlined briefly.

1.2. DNA methylation

5-methyl-dexycytidine (5mC) was described in DNA as early as in 1925 from the tubercle bacillus (Johnson and Coghill, 1925), and two decades later in calf thymus DNA (Hotchkiss, 1948). However, the biological function of this cytosine modification was not illuminated at the time. In 1975, two papers (Holliday and Pugh, 1975, Riggs, 1975) proposed 5mC as an important epigenetic modification that could influence gene expression and this change the scientific community's view on the importance of this 'fifth nucleotide' in eukaryotic biology (Doerfler, 2006). Now DNA methylation is generally agreed as a typical epigenetic mark, as it satisfies the stringent criterion of an epigenetic system, 'mitotically and meiotically heritable' proposed by Riggs (1996) (Russo et al., 1996), which was redefined

from the original term defined by Waddington (Waddington, 1942). In prokaryotes, methylation at both adenine and cytosine residues were found as part of the host restriction system that serves to protect the prokaryotic cell from foreign genetic material especially for the bacterial genome (Bickle and Kruger, 1993). However, in multicellular eukaryotes, DNA methylation occurs predominately but not exclusively at cytosine residues within a CpG dinucleotide (Bestor, 2000).

In mammals, a major form of epigenetic information is carried by DNA methylation, which comes in the forms of 5-methylcytosine (5mC) and the more recently discovered 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009, Kriaucionis and Heintz, 2009) (Figure. 1.1). For 5mC, a methyl group is added covalently to the 5-position of cytosine by DNA cytosine methyltransferases (DNMTs), mostly within the context of CpG dinucleotides in mouse somatic cells; however, non-CpG methylation also occurs at a high frequency in mouse and human embryonic stem (ES) cells (Lister et al., 2009, Ramsahoye et al., 2000). Interestingly, non-CpG methylation seems to be a feature of the pluripotent state, as it is present in induced pluripotent stem cells (iPS) generated by transduction of a non-pluripotent somatic cell with stem cell-associated genes, which results in reprogramming of the recipient cell's epigenetic profile (Lister et al., 2009). A recent discovery is the identification of a second modification in vertebrate DNA, 5-hydroxymethylcytosine (5hmC), in Purkinje neurons and embryonic stem cells (Tahiliani et al., 2009, Kriaucionis and Heintz, 2009). 5hmC is formed by adding a hydroxy group to the methyl-cytosine, in a reaction mediated by the Tet (Ten Eleven Translocation) family of enzymes (Tahiliani et al., 2009, Iyer et al., 2009) (Figure 1.1). Its importance in epigenetics is that the hydroxymethyl group is suggested to alter the biological properties of methylated DNA. It is noteworthy that the methyl groups of both thymine and 5mC are susceptible to oxidation with 5mC being slightly more reactive resulting in the generation of 5hmC (Burdzy et al., 2002). 5hmC also presents a new experimental problem, as conventional techniques (with the exception of a 5hmC specific antibody), cannot distinguish between 5mC and 5hmC in DNA (Nestor et al., 2010, Huang et al., 2010b). Until this apparent anomaly is properly resolved, there will be some uncertainty in existing DNA methylation databases. Recent technical developments can now distinguish prominent 5hmC sites in the genome (Song et al., 2011a, Nestor et al., 2012). However, it is clear that 5hmC is less abundant than 5mC, and the latter is still the most prominent modification in vertebrate DNA in many tissues (Wu and Zhang, 2011). 5mC values are stable at a typical value of around 4.5% of all cytosine in tissues, whereas 5hmC values vary significantly (Munzel et al., 2010). This suggests that 5hmC has a specific function that is not absolutely correlated with 5mC levels; initial analysis suggests that 5hmC

is predominantly associated with the gene bodies of highly expressed genes (Song et al., 2011a). The replacement of a 5mC residue with 5hmC can inhibit the binding of methyl-CpG binding proteins (MECPs), interfering with their functional roles in transcription (Jin et al., 2010). Moreover, in acute myeloid leukemia (AML), inhibition of TET function is associated with hypermethylation of a fraction of CpG islands (Figueroa et al., 2010).

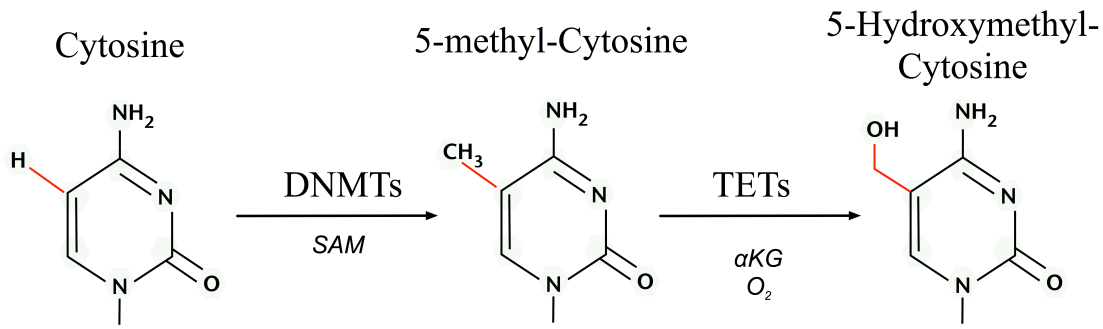


Figure. 1.1. The three major forms of cytosine bases in mammalian DNA. A methyl group can be added covalently to the 5-position of cytosine by DNA cytosine methyltransferases (DNMTs) with co-factor S-adenosyl methionine (SAM), resulting in the 5-methyl-cytosine, mostly within the context of CpG dinucleotides in somatic cells. In addition, 5-hydroxymethyl-cytosine (5hmC) is formed by adding a methyl group to cytosine and subsequently a hydroxy group, in a reaction mediated by the Tet (Ten Eleven Translocation) family of enzymes, α-keto gluterate (αKG) dioxygenases.

1.3. DNA methylation machinery

The presence of DNA methylation at regulatory sequences in somatic cells is generally associated with transcriptional repression and potentially has a long term impact on the stability of gene expression and on genome stability (Sharma et al., 2010). In humans, alterations in genomic methylation patterns are involved in the etiology of imprinting syndromes such as Beckwith-Wiedemann, Prader-Willi and Angelman, and have been implicated in a number of other disease conditions, especially cancers (Goll and Bestor, 2005). The presence of the methyl group to 5-position of the cytosine does not interfere with the Watson–Crick base paring of DNA, and this additional modification can be efficiently detected by DNA-interacting proteins (Jurkowska et al., 2010). In general, DNA methylation of the gene promoter region that is usually CG-rich results in the repression of the gene activity (Lister et al., 2009, Laurent et al., 2010, Zhang et al., 2009), while active genes usually correlate with hypomethylation around the regions of transcriptional start site (TSS) and high levels of methylation in the gene body (Lister et al., 2009, Laurent et al., 2010,

Zhang et al., 2009). In mammals, the enzymes responsible for targeting and maintaining global DNA methylation are constructed from a complex set of functional modules, broadly divided into the N-terminal 'regulatory' domain and the C-terminal 'catalytic' domain (Figure. 1.2). The regulatory domain functions largely as an interaction module, allowing multiple protein-protein interactions, DNA binding and nuclear localization (Goll and Bestor, 2005). Conversely, the C-terminal domain comprises ten motifs responsible for the enzyme's catalytic activity; six of these motifs are conserved in nearly all cytosine methyltransferases across the evolutionary spectrum from bacteria to mammals (Figure. 1.2).

Three methyltransferase enzymes, DNMT1, DNMT3A and DNMT3B coordinate the establishment and maintenance of DNA methylation patterns in mammals (Figure 1.2). The 'de novo' methyltransferases, DNMT3A and DNMT3B, target cytosine methylation to previously unmethylated CpG dinucleotides, whereas the 'maintenance' enzyme, DNA methyltransferase-1 (DNMT1), preserves existing methylated sites (Goll and Bestor, 2005). DNMT3A and 3B are thought to be de novo methylases with an equal preference for hemimethylated and unmethylated DNA, which are necessary for de novo methylation of the genome during development and potentially newly integrated retroviral sequences (Okano et al., 1999, Okano et al., 1998). The N-terminal region interacts with many chromatin-associated proteins including the de novo methyltransferases, methyl-CpG binding proteins (MECPs) and histone modifying enzymes. It also contains a nuclear localization signal, a PCNA (proliferating cell nuclear antigen)-interacting domain, a replication-targeting region and a cysteine-rich Zn^{2+}-binding domain that can potentially bind non-methylated CG rich DNA. DNMT1 also contains a domain showing homology to the polybromo-1 protein and is thought to mediate protein–protein interactions. Many of these interactions are involved in transcriptional repression (Smallwood et al., 2007, Esteve et al., 2006). Recent structural analysis suggests an elegant model for the maintenance DNA methylation function of DNMT1 in which its CXXC domain specifically binds to unmethylated CpG containing DNA and resulting in a repositioning of the CXXC-BAH1 linker between the DNA and the active site of DNMT1, preventing de novo methylation (Song et al., 2011b). Furthermore, a loop projecting from the BAH2 domain interacts with the target recognition domain (TRD), stabilizing it in a retracted position, and preventing it from accessing the DNA major groove. In consequence only hemimethylated CpG dinucleotides that do not bind the CXXC domain can gain access to the active site. The multiple interactions of DNMT1 suggest that it can be a participant in multiple complex networks involved in gene regulation, epigenetic signalling and genome stability. DNMT1 is also post-translationally modified and this can modulate its

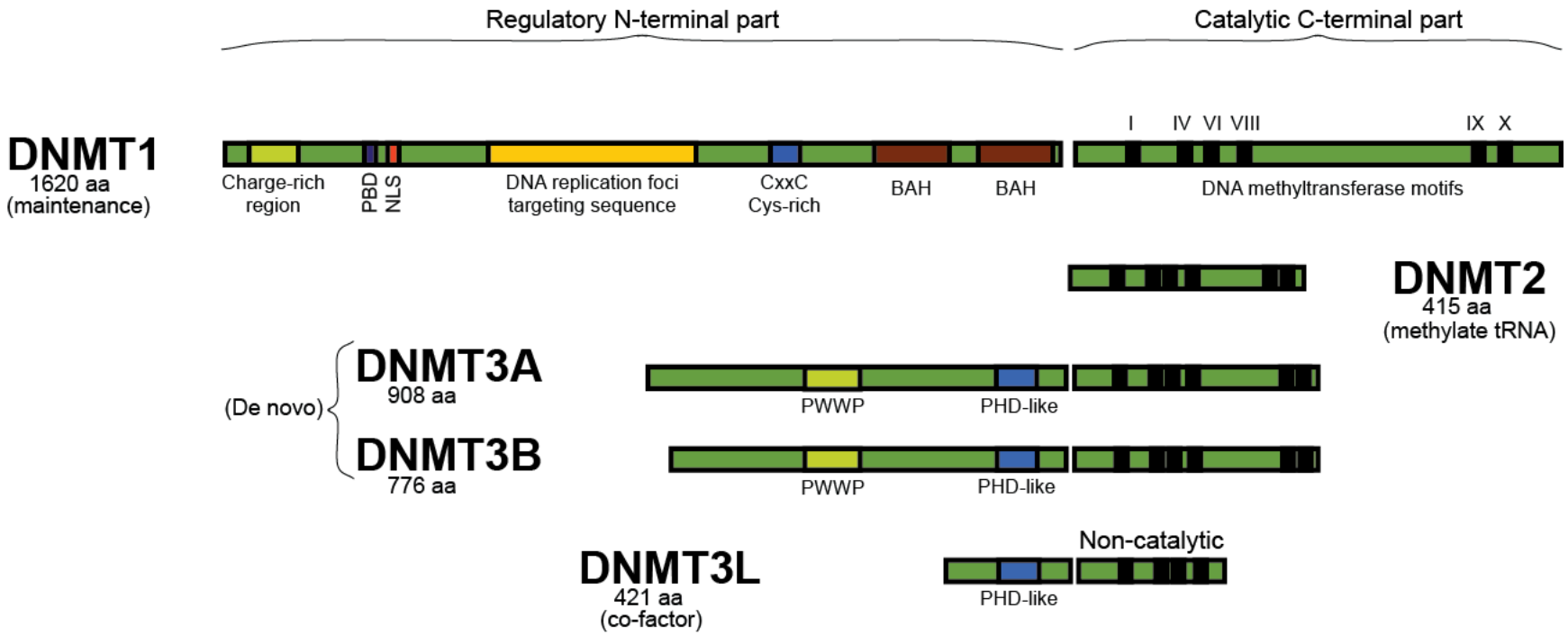


Figure. 1.2. Domain structures of mammalian DNA methyltransferases (DNMTs). Functional domains in the N-terminal part of the DNMTs are shown and the conserved motifs in the C-terminal part are labelled. In the N-terminal part the sub-domains labelled are: PCNA binding site (PBD), nuclear localization signal region (NLS), bromo-adjacent homology domains (BAH), plant homeo domain (PHD) like domain, and PWWP domain (highly conserved proline- tryptophan-tryptophan-proline motif, involved in protein–protein interactions). *N*- and C-terminal domains are linked with Gly-Lys dipeptides. In the C-terminal part the highly conserved methyltransferase motifs are shown as black thick lines (indicated as I-X).

protein-protein interactions, protein-DNA interactions, subcellular localization, catalytic activity and its stability (Patel et al., 2010, Du et al., 2010b, Lee and Muller, 2009). The protein lysine methyltransferase SET7 regulates DNMT1 activity in mammalian cells by promoting degradation of DNMT1 and thus allows epigenetic changes via DNA demethylation (Esteve et al., 2009). The Lys-142 of DNMT1 can be specifically monomethylated by SET7 and this modification is mutually exclusive with its Ser-143 phosphorylation; phosphorylated DNMT1 is more stable than methylated DNMT1 (Esteve et al., 2011).

The respective N-terminal domains of DNMT3A and 3B are responsible for their targeting to chromatin via a PWWP domain, and impart differences in specificity. In this respect, it appears DNMT3B is specialised in methylation of specific regions of the genome, such as pericentromeric repeats and CpG islands on the inactive X-chromosome, whereas DNMT3A is required for maternal imprints of differentially methylated regions (DMRs), in addition to their general de novo roles (Kim et al., 2009a). The PWWP domain of DNMT3A specifically recognizes the histone 3 lysine 36 trimethylation mark and this may be important for its subnuclear localization (Dhayalan et al., 2010). This interaction may underlie co-distribution of DNA methylation and H3K36me3 in chromatin (Hodges et al., 2009). Mouse embryos become hypomethylated during early development; reestablishment of global methylation depends on the de novo activity of DNMT3A and DNMT3B (Okano et al., 1999). Once established, DNA methylation patterns are stably maintained over cell divisions by DNMT1 through its preference for hemimethylated DNA. In addition, deletion of DNMT3A in primordial germ cells disrupts paternal and maternal imprinting, where as DNMT3B is dispensable for mouse gametogenesis and imprinting (Kaneda et al., 2010, Kato et al., 2007). A rare chromosome breakage disease called the immunodeficiency, centromeric region instability, and facial anomalies syndrome (ICF) is associated with point mutations in the DNMT3B gene leading to hypomethylation of satellite DNA, this hypomethylation phenotype can be recapitulated in a mouse model (Velasco et al., 2010, Xu et al., 1999).

Protein interaction domains in the regulatory N-terminal of DNMT3A and DNMT3B also mediate binding to transcriptional co-repressors (Qiu et al., 2002). DNMT3B is associated with SIN3A, SUMO1/UBC9, condensin and the chromatin remodeling enzyme hSNF2H, while DNMT3A has been shown to bind the co-repressor RP58 and the oncogenic factor PML-RAR (Goll and Bestor, 2005). Both also interact with DNMT1, histone deacetylases (HDAC) and multiple transcription factors (Hervouet et al., 2009). Thus, the cooperative function between DNMTs provides a way of passing on and maintaining

epigenetic information between successive cell generations and this is reinforced through interaction with chromatin modifiers. Unlike DNMT1 and DNMT3A/B, the DNA methyltransferase DNMT2 has only weak methyltransferase activity in vitro, and its absence causes no discernable effects in global CpG methylation levels nor results in a developmental phenotype (Goll and Bestor, 2005). A cofactor, DNMT3L (DNMT3-Like), is expressed only in germ and ES cells. It is not a methyltransferase but enhances de novo methyl transferase activity in ES cells (Ooi et al., 2010).

In addition to a role in gene silencing, 5mC is a prominent source of mutations as its mutation to thymine (T) is 10–50-fold higher than other transitions (Duncan and Miller, 1980). Both DNMT3A and 3B can interact with repair enzymes, thymine DNA glycosylase (TDG) and methyl-CpG binding domain protein 4 (MBD4), that have evolved to counteract the mutagenic effect of methylcytosines (Boland and Christman, 2008, Li et al., 2007, Hendrich et al., 1999). This implies a mechanistic link between DNA repair at sites affected by 5mC deamination and subsequent re-methylation.

1.4. DNA methylation, RNA interference, histone modifications and repression

The presence of DNA methylation (5mC) at CG rich promoters clearly acts as a deterrent to transcription. Known targets include imprinted genes, transposons, the XIST gene and CpG islands (CGIs) genes on the inactive X-chromosome. However, the extent to which promoter methylation regulates developmental and tissue specific gene expression is unresolved. In general, repression by DNA methylation is considered to occur downstream of other epigenetic or trans-acting factors that signal the initial inactivation event. For example, in the case of the pluripotency factor POU5F1 (previously known as OCT3/4), initial repression during differentiation is mediated by sequence-specific repressors such as GCNF followed by conversion from an active to an inactive histone modification signature and lastly by the deposition of DNA methylation at the promoter (Cedar and Bergman, 2009, Feldman et al., 2006). Genome wide sequencing analysis suggests that the number of potential genes that can be directly regulated by DNA methylation in a tissue and developmental specific manner may be quite small corresponding to 100–200 of annotated CGI's in somatic cells (Meissner et al., 2008). However, this may be an underestimate, as new data suggests there are approximately 23,000 and 25,500 CGIs in the mouse and human genomes respectively, about half of which are associated with annotated transcription start sites for mainly constitutively expressed

genes (Illingworth et al., 2010) (Figure. 1.3). As with previous estimates, around 2.5% of the annotated CGI's are methylated in somatic tissues and may be directly regulated by DNA methylation. In contrast the non-annotated or 'orphan' CGI's show higher levels of tissue specific methylation (14-20%) and lack marks of active transcription; histone H3 tri-methylation on lysine 4 (H3K4me3) and RNA polymerase II occupancy (Illingworth et al., 2010) (Figure. 1.3). These may be directly regulated by DNA methylation in different tissues and developmental stages. The importance of the preservation of these patterns is suggested by the abnormal silencing of annotated tumor suppressor genes by unscheduled de novo methylation of their promoter CGIs, which can contribute to unchecked proliferation in neoplastic cells (Sharma et al., 2010). In addition, tumor-specific CGI methylation differed from that in normal tissues by not being preferentially targeted to orphans CGIs (Illingworth et al., 2010). At the same time as de novo methylation of CGIs, global methylation levels associated with satellite repeats and retroposons are reduced in cancers. This may permit activation of typically silent transposons and contribute to genomic instability through illegitimate recombination events (Sharma et al., 2010).

RNA interference is a conserved process by which sequence-specific long double-stranded RNA (dsRNA) is converted into short fragments of ~20 nucleotides that are called siRNAs (small interfering RNA), which can induce gene silencing via two pathways: post-transcriptional gene silencing and transcriptional gene silencing (TGS) (Hamilton and Baulcombe 1999, Matzke et al., 2001). In plants, RNNA-directed DNA methylation (RdDM) has been shown by genetic and biochemical approaches, in which sequence-specific DNA methylation can be established by small interfering RNAs (siRNAs) guided de novo DNA methyltransferases (Chinnusamy and Zhu, 2009, Law and Jacobsen, 2010, Matzke et al., 2007, Wassenegger et al., 1994, Zhang and Zhu, 2011). TGS is well documented in plants but less so in mammals (Matzke et al., 2001). It has been demonstrated that promoter-targeted siRNAs can induce silencing of simian immunodeficiency virus (SIV) replication by induction of methylation at a CpG site within the SIV promoter region following siRNA-induced suppression (Lim et al., 2008, Suzuki et al., 2005). Different classes of small RNAs differ in their origin, biogenesis, expression pattern, and utilization. MicroRNAs (miRNA) are single-stranded RNA molecules approximately 21–23 nucleotides in length that can silence by binding to target mRNAs. It has been suggested that miRNAs are also directly involved in the maintenance of genomic integrity through global repression of transposable elements (TEs) and indirectly through regulation of DNA and histone modifying enzymes (Shalgi et al., 2010, Das et al., 2010). Interestingly, a noncoding RNA has been identified that interacts with the ribosomal DNA promoter in mouse NIH3T3 cells that mediates

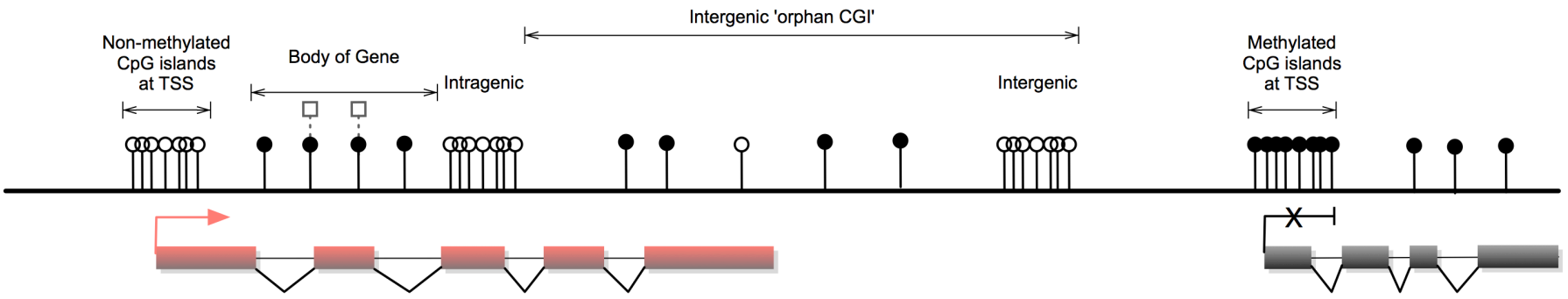


Figure. 1.3. The distribution of CpG dinucleotides in mammalian genomes. In vertebrate genomes, CpG dinucleotides are generally highly methylated with the punctuations of so-called CpG islands (CGIs), which are highly CG regions that are methylation-free and associated with promoters. The exceptions are those CGIs on the inactive X chromosome in female cells, where CpG islands are hypermethylated. In addition to the canonical CGIs location at annotated transcription start sites (TSSs), the CGIs within gene bodies (intragenic) and between annotated genes (intergenic) of unknown function were identified and named as 'orphan CpG islands'. Unmethylated CpG islands at 5' end of many genes positively correlate with transcriptional activity (active, left), whereas a small number of gene in specific cell types are hyper-methylated at their promoter CpG islands, which associates with transcriptional repression (inactive, right). Gene bodies are often methylated, with higher DNA methylation at their exons while lower at introns. Additionally, 5-hydroxymethylcytosines (5hmC) are also present at expressed gene bodies, which are proposed as the oxdation products of 5mC catalysed by Tet proteins (labelled white squares at body of gene). White circles: non-methylated CpGs; Black circles: methylated CpGs. Grey squares: hydroxyl-methylated CpG. Red boxes: active and transcripted exons; black boxes: inactive and silenced exons; transcriptional states of these genes are represented by the red arrow (active) or the black cross (inactive).

recruitment of DNMT3B and subsequent silencing of ribosomal RNA genes (Schmitz et al., 2010).

While DNA methylation plays an important part in overall transcriptional repression, it is clear that animal cells also utilize additional networks to mediate gene silencing. In ES cells that lack all three active DNA methyltransferases, a relatively small percentage of genes become reactivated (Fouse et al., 2008). This contrasts with inhibition of Dnmt1 function in somatic cell (Jackson-Grusby et al., 2001). Eukaryotic DNA is packaged with histone and non-histone proteins into chromatin and this has two important consequences. Firstly, it allows the packaging and compaction of ~1.8 m of DNA into a nucleus typically 5-20 μm in diameter. Secondly, it promotes the regulation of essential cellular events including transcription, lineage specification, DNA replication and cell division. The fundamental unit of chromatin is the nucleosome, which is composed of an octamer of the four core histones (H2A, H2B, H3 and H4) around which 147 bp of DNA is super-helically coiled 1.65 times (Khorasanizadeh, 2004). The repressive effect of nucleosomes on transcription can be enhanced or reduced by combinations of histone modifications (Lee et al., 2010, Shilatifard, 2006). The major effect of histone modifications is to regulate transcription by creating a docking site for non-histone effector proteins to modify chromatin structure. However in some cases, strongly charged modifications can alter chromatin structure directly, through disrupting the DNA-histone interaction. A diverse range of histone modifications have been reported including: acetylation, methylation, phosphorylation, ubiquitination, sumoylation, formylation, deimination, ADP ribosylation, and proline isomerization. Covalent post-translational modifications to the tails of histone proteins can reversibly affect gene expression by modifying their interaction with DNA and other nuclear proteins. These modifications can have either activating or repressive effects on the expression of surrounding genes, depending upon which histone residue receives the particular modification. The combination of these modifications and the resulting effect on gene expression is referred to as the "Histone Code" (Turner, 2007).

For example, trimethylation of lysine 4 on histone H3 (H3K4me3) is enriched at transcriptionally active gene promoters, whereas trimethylation of H3K9 (H3K9me3) and H3K27 (H3K27me3) is present at gene promoters that are transcriptionally repressed. The latter two modifications together constitute the two main silencing mechanisms in mammalian cells, H3K9me3 working in concert with DNA methylation and H3K27me3 largely working exclusive of DNA methylation. Genome-wide studies showing distinct localization and combinatorial patterns of these histone marks in the genome have significantly increased our understanding of how these diverse modifications act in a

cooperative manner to regulate global gene expression patterns (Sharma et al., 2010). The polycomb complex (PRC), which mediates tri-methylation of lysine 27 on histone H3 (H3K27me3) appears to be targeted specifically to genes involved in development and differentiation (Mikkelsen et al., 2007). DNA methylation and specific histone modifications can also influence each other during mammalian development and it is noteworthy that many of the components from each system interact such that histone H3 methylation at lysine 9 (H3K9me2/3) can help to direct DNA methylation patterns, and DNA methylation might serve as a template for some histone modifications after DNA replication (Lee et al., 2010, Cedar and Bergman, 2009). Heterochromatin protein (HP1) binds H3K9me2/3 containing chromatin through its chromodomain and at the same time acts like a nuclear 'Velcro' protein through multiple interactions with protein partners via its chromoshadow domain (Dialynas et al., 2008). H3K9me2/3 is generated by a multitude of histone methyltransferases including SUVAR39H1, G9a and GLP. Interaction partners for HP1 include DNMT1, SUVAR39H1 and G9a (Smallwood et al., 2007, Esteve et al., 2006). HP1 can inhibit preinitiation complex (PIC) assembly in vitro by blocking key subunits of the TFIID and Mediator coactivator complexes. Notably, binding of HP1 inhibited the Sp1-regulated survivin gene in vivo upon DNA damage-induced silencing (Smallwood et al., 2008). Loss of HP1 proteins causes chromosome segregation defects and lethality in some organisms and a reduction in levels of HP1 family members is associated with cancer progression in humans (Dialynas et al., 2008).

In contrast, histone H3 methylation at lysine 27 may be mutually exclusive with the presence of DNA methylation (Weber et al., 2007). Genes targeted by polycomb actually undergo repression through a process of heterochromatinization. Binding of the PRC2 complex to specific genes brings about local tri-methylation of histone H3K27 by means of the histone methyltransferase EZH2 contained in these complexes (Morey and Helin, 2010). These methyl groups then serve as a ligand for the chromodomain protein, CBX2, that is part of the PRC1 complex, and this generates a heterochromatin-like structure and gene repression. However polycomb target genes in ES cells can have a bivalent chromatin signature, being marked by both H3K27me3 and the activating modification H3K4me3 (Mikkelsen et al., 2007, Barski et al., 2007). As development proceeds, the polycomb complex is removed in a gene and cell-type specific manner, thus activating or maintaining the silencing of those genes as required through development (Mikkelsen et al., 2007, Boyer et al., 2006). Like DNA methylation, gene silencing via histone modification can be maintained in vivo through multiple cell divisions. One possibility is that this occurs through the simple rebinding of repressor molecules following DNA replication.

It has been reported that CGIs that are aberrantly methylated in neoplastic cells coincide with sites targeted by polycomb in human ES cells (Schlesinger et al., 2007). Approximately 16% of all CGIs are H3K27me3 positive in human ES cells and these sites are not over-represented among the CGIs methylated in blood, cerebellum or normal colon (Illingworth et al., 2010). In contrast, 56% of tumor-specifically methylated CGIs are derived from CGIs that were H3K27 trimethylated in embryonic cells (Illingworth et al., 2010). These findings emphasize the distinction between tumor-specific CGI methylation and that found in normally developing human somatic tissues and reinforce the possible link between polycomb complexes and tumor-specific CGI methylation. Loss of the Swi/Snf chromatin remodeler component SNF5 in malignant rhabdoid tumors leads to elevated expression of EZH2 (Wilson et al., 2010). Consequently, Polycomb targets are broadly H3K27-trimethylated and repressed in SNF5-deficient fibroblasts and cancers. Normally there is antagonism between SNF5 and EZH2 in the regulation of stem cell-associated programs so that SNF5 loss deregulates these programs.

1.5. Methyl-CpG binding proteins - interpreters of DNA methylation

The methylation mark can be specifically recognised by a group of methyl-CpG binding proteins (MECPs), which may mediate an alternative mechanism for transcriptional regulation. In general, the MECPs may 'read' the established methylated sequences, recruit histone modifying complexes, together with which regulate higher order chromatin structure, maintain genome integrity, and stabilize patterns of gene expression. At present three major families of mammalian MECPs have been characterized (Ballestar and Wolffe, 2001) (Figure. 1.4). 1. Methyl-CpG binding domain proteins (MBDs) including MBD1, MBD2, MBD3, MBD4 and MeCP2; 2. The structurally unrelated methyl-CpG binding zinc-finger proteins of Kaiso family including KAISO/ZBTB33, ZBTB4 and ZBTB38; and 3. Methyl-CpG binding SRA domain proteins of UHRF family including UHRF1 and very similar UHRF2. The first family of MBDs, with the exception of MBD3, all specifically recognize methyl-CpG through their novel methyl-CpG binding domain (MBD). In contrast, KAISO family proteins bind methylated DNA through a zinc finger motif. The third family of UHRF proteins recognize methylated DNA via a SRA domain and contain an ubiquitin-like motif, a tudor domain, a PHD domain, and a ring finger (Figure. 1.4).

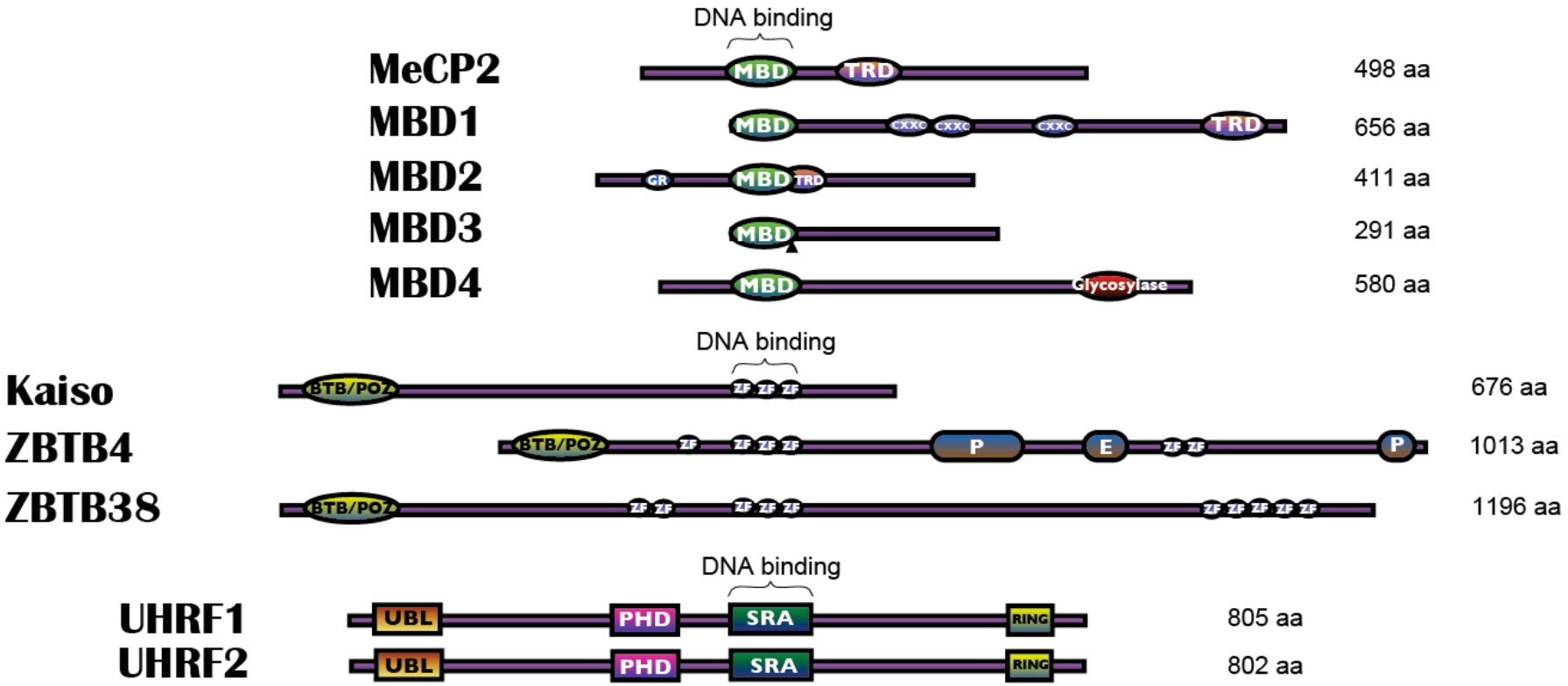


Figure. 1.4. Domain structures of methyl-CpG binding proteins. At present three families of mammalian MeCPs have been characterized: 1. The methyl-CpG binding domain proteins (MBDs) including MBD1, 2, 3, 4 and MeCP2; 2. The structurally unrelated methyl-CpG binding zinc-finger proteins of Kaiso family include KAISO/ZBTB33, ZBTB4 and ZBTB38; 3. The methyl-CpG binding SRA domain proteins of UHRF family including UHRF1 and very similar UHRF2. The sub-domains labeled are MBD: methyl-CpG binding domain; TRD:trans-repressor domain; GR, E, P: amino acid repeats; BTB/POZ: broad complex, tramtrack, and bric à brac domain; ZF: zinc finger motis; UBL: ubiquitin-like motif; PHD: Plant-and homeodomain; SRA: SET and Ring- associated domain. The DNA binding regions are indicated.

MECPs have binding affinity specifically to symmetrical 5-methyl-CpG (5meCpG) but some of them have additional binding preferences. The MBD domain of MBD1 recognizes methylated CpGs within TCGCA and TGCGCA sequence context more efficiently and CXXC motifs of MBD1 can recognise and bind to unmethylated DNA (Clouaire et al., 2010). The MBD of chicken MBD2 (96% identical to human MBD2) is shown to bound to methyl-CpG in a single orientation, which suggests its sequence specificity for bases outside of the symmetric methyl-CpG (Scarsdale et al., 2011). MeCP2 requires an A/T run adjacent to the methyl-CpG for efficient DNA binding, which were shown by in vitro binding assays (Klose et al., 2005). Mammalian MBD3 contains a well-conserved MBD domain but it does not recognize methylated DNA, owing to amino acid substitutions within its MBD domain. Instead, MBD3 is an essential subunit of the Mi-2/NuRD chromatin remodelling complex, which have been shown to function as a transcriptional repressor complex both in vitro and in cell culture assays. In addition, two proteins MBD5 and MBD6 were identified based on their domain similarity to the MBDs, but no methylated DNA binding affinity was demonstrated (Laget et al., 2010). In the second MECP family, the zinc finger domain of KAISO proteins is responsible for the methyl-CpG binding, and has a binding preference to two methyl-CpG motifs in close proximity, preferably in tandem (Filion et al., 2006, Prokhortchouk et al., 2001, Ruzov et al., 2004). The third MECP family proteins UHRF proteins recognise methylated DNA through their SRA domain. UHRF1 is reported to bind hemi-methylated DNA at symmetrical methylated CpGs. By a distinct flipping mechanism similar to that of DNMT1, the SRA domain of UHRF1 can recognise and bind to hemi-methylated DNA, and is involved in the maintenance of DNA methylation with DNMT1 (Avvakumov et al., 2008, Arita et al., 2008). The other protein in the family UHRF2 has an SRA domain with 75% similarity to that of UHRF1 and has been shown to have a similar binding preference to hemi-methylated DNA in vitro (Pichler et al., 2011). As there are a large number of proteins containing zinc finger or SRA domains in the genome, it is very possible more methyl-CpG binding proteins may be identified with distinct binding specificities.

Almost all the MECPs have been demonstrated to be associated with transcriptional repressors and induce transcriptional repression, suggesting an elaborate link between DNA methylation and transcriptional repression. In addition to the direct mechanism in which DNA methylation alters the transcription binding sites, interfering the binding of transcriptional factors such as E2F or CREB, thereby preventing transcriptional activation (Campanero et al., 2000, Iguchi-Ariga and Schaffner, 1989), MECPs can read and interpret the epigenetic mark of methylation and associate with various chromatin modifiers to

establish a repressive chromatin environment (Fuks et al., 2003, Jones et al., 1998, Nan et al., 1998, Wade et al., 1999). One of the examples is the founding MECP protein MeCP2; In addition to the binding specificity of methylated DNA, MeCP2 has been demonstrated to associate with various co-repressor complexes such as Sin3A, NCoR, through its transcription-repression domain (TRD) (Kokura et al., 2001, Jones et al., 1998). MeCP2 was demonstrated to target to promoter DNA and cause strong transcriptional repression, and inhibition of HDACs in the Sin3A complex by trichostatin A (TSA) relieves MeCP2-mediated silencing, suggesting that MeCP2 may serve as a global transcriptional silencer (Nan et al., 1998, Jones et al., 1998). However, only subtle changes in gene expression were identified by the global transcriptional profiling of $MeCP2^{-/-}$ mice brains (Tudor et al., 2002), suggesting other MECPs may compensate the functional defects caused by the lacking of MeCP2 activity. In response to this hypothesis, mice lacking three of the MECPs MeCP2, MBD2 and KAISO have been generated (Martin Caballero et al., 2009). All of these three MeCPs have been shown to be involved in the transcriptional repression. The phenotype of these mice is still viable and fertile with neurological abnormalities, which is very similar to mice lacking MeCP2 alone. It is noteworthy that the closest relative protein of MeCP2 within the MECPs, MBD4, was not included in this multiple knockout, as MBD4 was demonstrated mainly as a repair protein in the group at the time (Hendrich and Tweedie, 2003). In addition, there might be more MECP proteins which have not been identified so far, thus the mild phenotype of the triple-knockout mice only suggests that the chosen three proteins do not compensate for each other, in terms of their activities of transcriptional repression (Bartke et al., 2010). In addition, the MeCP2 interaction with the Sin3A chromatin remodelling complex is not stable, suggesting MeCP2 may engage a more diverse range of cofactors for its repressive function (Klose and Bird, 2004). Indeed, MeCP2 has been demonstrated to interact with other chromatin-associated complexes such as nuclear co-repressor complex N-CoR (Kokura et al., 2001), and the core ATPase of the SWI/SNF ATP-dependent chromatin remodeling complex, Brm (Hu et al., 2006, Harikrishnan et al., 2005). The functional role of these associations is not clear yet, but interacting convergence of these repressive complexes implicated from several other MECPs such as KAISO and MBD2 implicates the importance of these associations. Moreover, the MBD domain of MBD4 is most similar to that of MeCP2, with common MBD ancestor shared based on the evolutionary primary sequence analysis (Figure. 1.7). Due to the special glycosylase domain property of MBD4, it was thought mainly as a repair protein. The possible repressive function of MBD4 at its regulatory N-terminal may have been underestimated.

1.6. Cellular response to epigenetic perturbation

The utilization of epigenetic regulatory mechanisms in disease states can contribute to the molecular pathology by altering gene expression states that support progression and inhibit defense pathways such as tumor suppressor genes (Sharma et al., 2010). For example, epigenetic inactivation of hMLH1 is a major cause of microsatellite instability in sporadic colorectal cancers, likewise epigenetic alteration of genes involved in the induction of senescence is often associated with cancers showing mutations in the Ras signaling pathway (Toyota and Suzuki, 2010). Hypomethylation of short and long interspersed repetitive elements has been reported in cancer, and hypomethylation of the genome has been observed in ICF syndrome (Jin et al., 2010, Tao et al., 2002). Since epigenetic mechanisms are deeply embedded in regulating developmental gene expression programs, it is not surprising that loss of components of the epigenetic toolbox during development results in severely impaired or embryonic lethal phenotype. The cellular response is not governed just by transcriptome changes but also specific activation of signaling cascades that usually respond to DNA damage. Despite obvious differences in biology and early developmental strategies, inhibition of dnmt1 function in *Xenopus laevis* embryos shows remarkable similarity with mouse *Dnmt1*$^{-/-}$ mutants as indicated by the presence of axial defects, failure to form a neural tube, and improper patterning of the somites (Stancheva and Meehan, 2000, Lei et al., 1996, Li et al., 1992). Underlying the phenotype in both species is activation of p53-mediated apoptosis (Stancheva et al., 2001, Jackson-Grusby et al., 2001). One generalization from the analysis of DNMT1 depletion or inhibition is that it is not essential for the survival of embryonic cells during early cleavage stages. Instead the effect of disrupting DNMT1 function only becomes apparent during and past gastrulation when loss of DNMT1 provides a signal through p53, to initiate apoptosis. Although reduction of *Dnmt1* levels has been associated with the maintenance of pancreatic cell identity (Dhawan et al., 2011, Oghamian et al., 2011, Anderson et al., 2009), which possibly depends on an event of DNA methylation-mediated repression of pancreatic cell lineage determination genes (Dhawan et al., 2011), this does not preclude the possibility of an impaired apoptosis pathway through p53 (Ruzov et al., 2009b). Recently it has also been reported that mutations in *DNMT1* cause autosomal dominant cerebellar ataxia, deafness and narcolepsy (Winkelmann et al., 2012), but the underlying molecular mechanism(s) remain unidentified.

Like DNMTs, inactivation of many histone-modifying enzymes in ES cells does not result in cell lethality but is detrimental to embryo development (Tachibana et al., 2005, Tachibana et al., 2002, Peters et al., 2001). The H3K9 methyltransferases G9a, GLP and

Suvar39h1/2 are required for early embryo development, although mortality of G9a and GLP deficient embryos occurs earlier (by E8.5) than Suvar39h1/2 mutants (E14.5) (Tachibana et al., 2005, Tachibana et al., 2002, Peters et al., 2001). The G9a and GLP mutants are very similar and TUNEL assays indicate that abundant apoptosis occurs in *G9a*$^{-/-}$ embryos, whereas very few cells were positive in wild-type embryos (Tachibana et al., 2002). The mechanistic basis of this apoptotic activation is unclear, however *G9a*$^{-/-}$ derived MEFs are viable despite having an extensive gene mis-expression profile, which suggests there is a developmental specific checkpoint (as for DNMT1) that is activated upon G9a inactivation in early embryos (Sampath et al., 2007). By contrast, *Suv39h* double mutants exhibited genomic instability in a subpopulation of cells, which did not trigger extensive apoptosis (Peters et al., 2001). This suggests that repeat DNA associated Suv39h-dependent H3-K9 methylation is important for maintaining a stringent higher-order structure at pericentric heterochromatin, which is required to protect genomic stability. Absence of Suv39h HMTase does not trigger cell death pathways. Similarly loss of the Polycomb components has not been reported to trigger apoptosis (Morey and Helin, 2010). However, the S-adenosylhomocysteine hydrolase inhibitor 3-Deazaneplanocin A (DZNep) induces efficient apoptotic cell death in cancer cells but not in normal cells by depleting cellular levels of the PRC2 components EZH2, SUZ12, and EED (Tan et al., 2007). This reduced H3K27me2/3 levels but not H3K9me2/3. In this case preferential reactivation of a novel apoptosis effector, FBXO32, by DZNep, may be responsible for inducing apoptosis in the cancer context. This would be consistent with a cancer specific phenotype in which epigenetic mechanisms have been co-opted to selectively repress genes that are antagonistic to transformation either by inhibiting cellular proliferation and promoting either senescence or apoptosis (Wu et al., 2010).

A recent report suggests that DNA methyltransferases are essential for maintaining the memory of a genotoxic insult that persists after γ-radiation exposure (Rugo et al., 2011). Mouse embryonic stem cells exposed to γ-radiation harbour the effects of the insult for weeks and conditioned media from the progeny of exposed cells can induce DNA damage and homologous recombination in naive cells. Experiments in ES cells suggest that the molecular pathway that underlies the memory of the insult requires Dnmt1 and Dnmt3a but not Dnmt3b (Rugo et al., 2011). These exciting data suggests a potential molecular pathway for persistent bystander effects associated with exposing cells to of DNA-damaging agents.

Ubiquitin-like, containing PHD and RING finger domains, 1, also known as UHRF1 or Np95, binds methylated CpG through its SET and RING finger-associated (SRA) domain (Sharif et al., 2007). It localizes to replicating heterochromatin and is dependent on the

presence of hemi-methylated DNA (Figure. 1.6 B). Np95 forms complexes with Dnmt1 and is essential in vivo to maintain global and local DNA methylation. *Np95*$^{-/-}$ embryos exhibited growth retardation and various malformations such as neural tube closure defect, small branchial arches and failure in outgrowth of allantois. Like *Dnmt1*$^{-/-}$ embryos they exhibit excess apoptosis that may be independent of the transcriptional changes associated with DNA hypomethylation.

Recent work demonstrates that complete inactivation of DNMT1 function in human cancer cells results in cell death, but this decrease in viability occurs with minimal changes in global DNA methylation (Chen et al., 2007). This observation supports the hypothesis that DNMT1 possesses essential functions independent of its role as a maintenance methyltransferase and links its absence with activation of a cellular checkpoint response. Partial reduction in xDnmt1p levels without changes in DNA methylation levels is sufficient to activate a cell death program in *Xenopus* embryos (Dunican et al., 2008). This contrasts with *Dnmt1*$^{-/-}$ ES cells, which proliferate normally in culture unless they are induced to differentiate, but notably they exhibit high rates of micro-satellite instability and DNMT1 (but not the de novo methyltransferases) can be recruited to sites of DNA damage via PCNA (Mortusewicz et al., 2005, Guo et al., 2004). A possible mechanism for DNMT1 recruitment and signaling apoptosis is through the mismatch repair pathway (MMR). The methyl-CpG binding protein, MBD4 (MED1) has been shown to function as a thymine glycosylase and interacts with the MMR protein MLH1. An important function of MMR proteins is to sense DNA damage and mediate the decision to repair the lesion or to induce apoptosis in somatic cells (Palii et al., 2008). The levels of several MMR proteins are reduced in *Mbd4*$^{-/-}$ mouse embryonic fibroblasts, which can account for the diminished apoptotic response of these cells to DNA damaging agents (Sansom et al., 2003, Cortellino et al., 2003). In wild type cells, DNA-damage recognition by MMR factors is sufficient to trigger cell-cycle arrest and apoptosis through direct interaction with signaling kinases such as ATM, ATR, CHK1, CHK2, which ultimately activates p53 (Luo et al., 2004). MBD4 can interact directly with both DNMT1 and MLH1 leading to recruitment of all three components at DNA damage sites (Ruzov et al., 2009). The co-localization of DNMT1, MBD4 and MLH1 suggests that they may participate in a cellular checkpoint that monitors potential DNA hypomethylation events by detecting the presence or absence of the maintenance methyltransferase, perhaps at or adjacent to the replication fork. The recruitment of these components in response to localized DNA damage suggests that they can have a role in the cellular decision whether to repair the lesion or activate apoptosis (Ruzov et al., 2009). Double depletion experiments in *Xenopus laevis* suggest that both MBD4 and MLH1 are required for the embryonic lethal phenotype

of associated with DNMT1 depletion. Inhibition of p53, MBD4 or MLH1 alleviates phenotypic consequences DNMT1 inactivation in frogs. This suggests in this model system that it is the absence of DNMT1 at cellular checkpoint that triggers apoptosis by releasing MLH1/MBD4 to activate downstream DNA damage kinases such as ATM, which activate p53 and subsequent apoptosis. In agreement with this, over-expression of either MLH1 or MBD4 in MEFs resulted in activation of an apoptotic response and the glycosylase catalytic activity of MBD4 is not required for this (Ruzov et al., 2009b). In contrast, over-expression of MBD4 did not induce apoptosis in HCT116 cells, which lack functional MLH1, whereas over-expression of GFP-MLH1 did. This model suggests that the maintenance of genome integrity through DNA repair mechanisms includes preservation of epigenetic signatures, requiring the participation of DNA and histone modifying enzymes in repair pathways.

In general, the propagation and preservation of epigenetic signatures in development is essential for normal transcription programs. The consequences of loss of these systems can contribute to susceptibility to disease and increased phenotypic variation (Whitelaw et al., 2010, Sharma et al., 2010). Loss of this transcriptional 'dampening' system is detected by cellular surveillance systems that in many cases can result in activation of intrinsic and extrinsic apoptotic pathways if the epigenetic alterations are detrimental to cell viability.

1.7. Methyl-CpG binding domain protein 4

MBD4 is the only methyl-CpG binding DNA protein that associated with repair function within the MECPs group. It has two significant domains identified by homology and functional analysis: a highly conserved glycosylase domain (60% conservation between species) at the C-terminus, and a less well conserved methyl-CpG binding (MBD) domain (37% conservation between species) at the N-terminus (Bellacosa et al., 1999, Hendrich and Bird, 1998) (Figure. 1.5). The N-terminal MBD domain of MBD4 is able to discriminate between methylated and unmethylated double stranded DNA (Hendrich and Bird, 1998a). The C-terminal glycosylase domain of MBD4 was shown in vitro to recognise and repair the mismatched products from deamination including thymine and uracil (Hendrich and Bird, 1998, Wong et al., 2002). MBD4 is proposed to be involved in the repair of mismatches resulting from cytosine deamination. Spontaneous hydrolytic deamination of methylated cytosine causes C to T transitions at methylated CpG, and C to U mutation at non-methylated CpG. MBD4 was shown to excise and repair both T and U mutations at methylated and non-methylated CpGs via its glycosylase domain. The methyl-CpG binding domain of MBD4 binds preferentially to methylated CpG:TpG mismatches, the primary

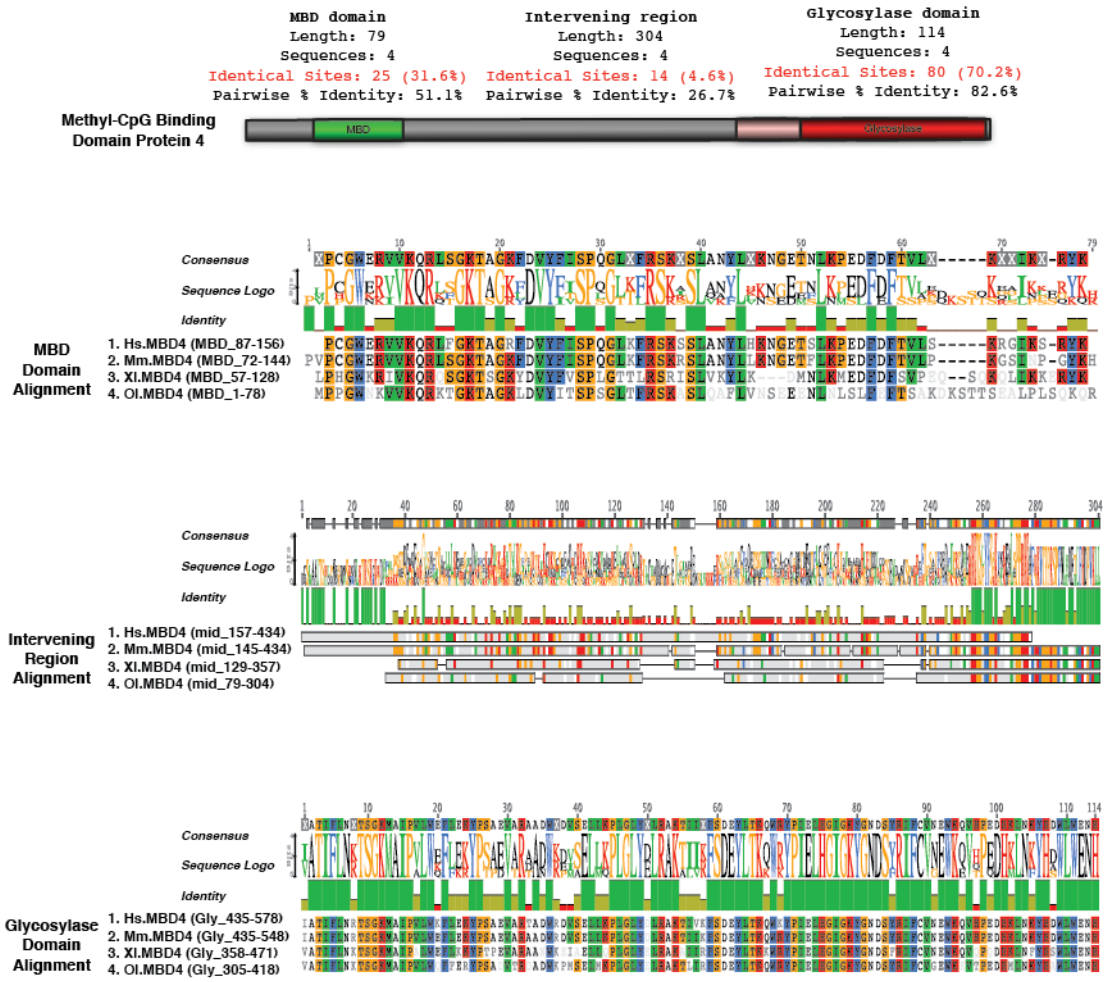


Figure. 1.5. Functional domains of methyl-CpG binding domain protein 4 (MBD4). A schematic highlighting the two domains of MBD4; an N-terminal MBD domain that binds methyl-CpG containing duplexes including mis-matches and a C-terminal Glycosylase domain, which can excise and repair both thymine and uracil mutations that result from deamination of methyl-cytosine and unmethylated cytosine. The glycosylase domain of MBD4 is well conserved, about 114 amino acids long, sharing 70.2% identical sites between species from Medaka Fish to Human. In contrast, the MBD domain of MBD4 has poor conservation, the 70 amino acids that constitute MBD domain sharing only 31.6% identical sites. The intervening region is a domain desert and highly variable in sequence. It is noteworthy that a 50 amino acid region right ahead of the glycosylase domain is conserved between species. This region interacts with several important protein partners, including MLH1 and FADD. Sequence alignments of the MBD domain, the intervening region and the glycosylase domain of MBD4 protein were generated with clustalX module of Geneious pro software.

product of deamination at methyl-CpG. MBD4 can repress several methylated reporter genes (Fukushige et al., 2006, Kondo et al., 2005), However, the global genomic targets of MBD4 have not been characterised. *Mbd4*$^{-/-}$ mice are both viable and fertile, and show only mild physiological defects (Millar et al., 2002). However, morpholino knock-down of xMbd4 in

Xenopus laevis shows serious developmental defects and fails to develop to tadpole. Re-introduction of mouse *Mbd4* mRNA can rescue the phenotype (Ruzov et al., 2009). Novel interacting partners of MBD4 include MLH1 and Fas-associated death domain (FADD) proteins, suggesting a potential link between genome surveillance and apoptosis (Screaton et al., 2003, Hendrich and Bird, 1998) (Figure. 1.6 D). Consistent with these observations, reduced apoptosis occurs in the small intestine of *Mbd4*$^{-/-}$ mice in response to a variety of DNA-damaging agents, and increased tumorigenicity was observed in *Mbd4*$^{-/-}$ mice with a tumor-susceptible *Apc* min background (Millar et al., 2002, Wong et al., 2002, Hendrich et al., 1999).

It has been proposed that MBD4 proteins arose as a fusion protein between MBD and glycosylase domain ancestors in the vertebrate lineage (Hendrich and Tweedie, 2003). MBD2/3 represents the ancestral methyl-CpG binding protein. Interestingly, a prototypical MBD4 protein and its putative ancestor MBD2/3 were both identified outside vertebrates in the cephalochordate amphioxus (*Branchiostoma floridae*) (Albalat, 2008). A putative MBD4 was predicted in the *Ciona intestinalis* genome, but it lacks an MBD domain. The finding of a putative MBD4 protein in this invertebrate-vertebrate transition model organism pushes back the origin of MBD4 proteins in evolutionary time (Albalat, 2008a). The MBD domain of MBD4 is most similar to that of MeCP2 in primary sequence, binds specifically to methylated DNA (Hendrich and Bird, 1998, Ballestar and Wolffe, 2001). MeCP2 has been described in several molecular roles including transcriptional repression, activation of transcription, nuclear organization, and splicing, whereas the role of MBD4 remains largely unknown apart from functions in DNA repair and apoptosis (Sasai and Defossez, 2009). Noticeably, the primary sequence of the MBD domain of MBD4 proteins in vertebrates diverges much more than that of MeCP2, while important residues responsible for DNA binding and Rett syndrome mutations are well conserved (Chapter 6).

MBD4 (NM_003925) is a gene of 9060bp coded by 8 exons from 129158852 to 129149792 according to GRCh37 human genome assembly v37 by IGB browser. The longest human MBD4 protein has 580 amino acids, at approximately 66 kDa. MBD4 gene mutations were found in tumors that exhibit genomic instability associated with defective DNA mismatch repair (MMR); termed primary microsatellite-instability (MSI). A characteristic of MSI tumors is frequent silencing of MLH1 and down-regulation of MMR target genes, such as MRE11 and MBD4 (Miquel et al., 2007). The somatic MBD4 mutations in these tissues are likely a consequence of MMR deficiency. Accumulated evidences have shown that MBD4 and several MMR proteins fall in the same base excision repair (BER) pathway. Interestingly, steady state amounts of several MMR proteins were found to be downregulated

Maintainance of DNA Methylation

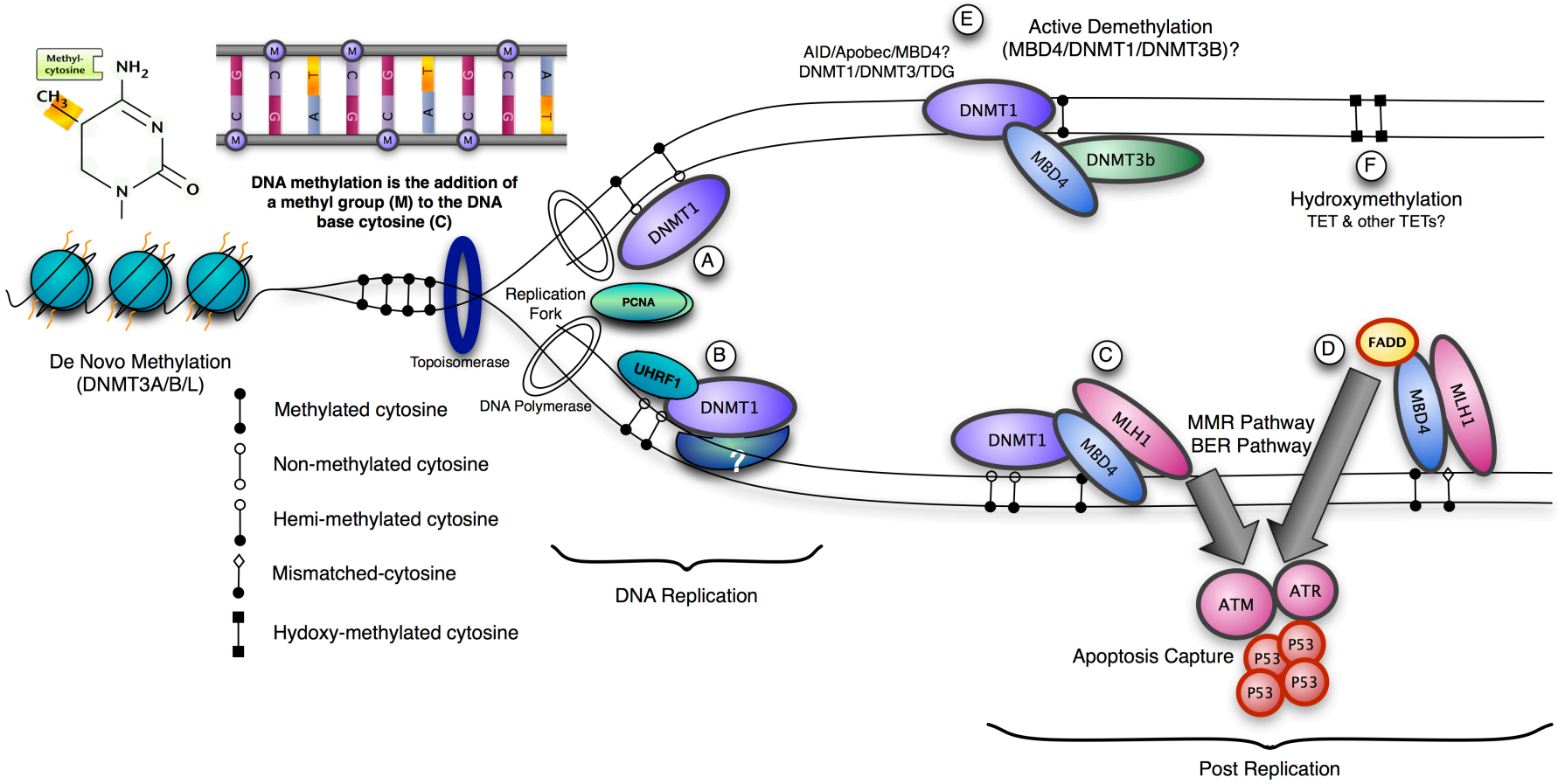


Figure. 1.6. Model illustration of the maintenance of DNA methylation and the involvement of protein complexes. (A-B) The preferential substrate for the maintenance cytosine methyltransferase, DNMT1, is hemi-methylated CpG sites resulting from newly synthesized DNA in somatic cells. DNMT1 can be localised at the replication fork, and functions with the help of partner proteins including UHRF1 and PCNA (Proliferating cell nuclear antigen). (C) At post-replication, DNMT1 and a methyl-CpG binding protein MBD4 can be localized together at DNA damage sites and may be part of cellular pathway that activates apoptosis. MBD4 interacts directly with both DNMT1 and MLH1 leading to recruitment of all three at DNA damage sites. (D) MBD4 has also been shown to recruit Fas-associated death domain protein (FADD), which bridges death receptors with initiator caspases. FADD may also be an apoptotic effector via MBD4. (E) Several active DNA demethylation models have been proposed. MBD4 was reported to execute active DNA demethylation at the CYP27B1 promoter in response to PTH (Parathyroid Hormone) signaling. Similarly TDG (Thymine DNA glycosylase) can also interact with DNMT3A and DNMT3B and function as 5mC glycosylase against hemi-methylated DNA with the same weak excision activity as MBD4. In zebrafish embryos, Aid, Mbd4 and the DNA repair protein Gadd45a may cooperate to induce demethylation. Thymine glycosylases such as TDG and MBD4 may function on deamination of 5-methyl-cytosine by repairing the resulting mismatch. (F) TET1 is capable of acting on both fully methylated and hemi-methylated DNA, producing 5-hydroxymethylcytosine (5hmC) in DNA, which may also act in signaling pathways associated with turnover and maintenance of the epigenome.

in *Mbd4*$^{-/-}$ MEFs (Cortellino et al., 2003). Although Mbd4 inactivation did not increase mini-satellite instability (MSI) in the mouse genome, it did result in a 2- to 3-fold increase in C to T transition mutations at CpG sequences in splenocytes and epithelial cells of the small intestinal mucosa (Millar et al., 2002). The change in the gastrointestinal cancer phenotype was associated with an increase in somatic C to T mutations at CpG sites within the coding region of the wild-type *Apc* allele. These studies indicate that *Mbd4* inactivation may alter the mutation spectrum in cancer cells and modify the cancer predisposition phenotype. A C-terminal truncation of MBD4 containing its MBD domain was detected in the HCA7 cell line (colon cancer cell line) (Bader et al., 1999). Frequent MBD4 mutations were reported in MSI tumours in sporadic colon cancers and in hereditary non-polyposis colorectal cancer (HNPCC) tumours. The frequency of MBD4 mutations was correlated with MSH2/MLH1 germline mutations (Riccio et al., 1999). In human, frame-shift mutation of MBD4 occurs in up to 43% of microsatellite unstable colon cancers. A frame-shift in a polyadenine tract due to an MMR defective background resulted in a mutated form of MBD4 that occurs naturally in human colon cancers (Abdel-Rahman et al., 2008). This leads to a premature truncation of the MBD4 protein, which has an intact MBD domain but lacks of the whole glycosylase domain. It has been suggested that this truncated form of MBD4 acts in a dominant negative way, competitively inhibiting normal glycosylase activity of wild type MBD4, and increases

mutation frequency when overexpressed in cells (Bader et al., 2007). The frameshift mutations in MBD4 appear to be selected for in MSI tumour cells, as similar coding polyadenine repeats in other tested genes are rarely altered in this panel of 56 primary MSI tumours (Riccio et al., 1999). Over-expression of MBD4tru in Big Blue (lacI)-transfected, MSI human colorectal carcinoma cells doubled the mutation frequency, indicating that the modest dominant negative effect on DNA repair can occur in living cells in short-term experiments. Intriguingly, the whole mutation spectrum was increased, not only at CpG sites, suggesting that truncated MBD4 has a more widespread effect on genomic stability. This demonstration of a dominant negative effect may be of significance in tumour progression and acquisition of drug resistance (Abdel-Rahman et al., 2008, Bader et al., 2007). Regarding tumor prognosis, reduced expression of MBD4 correlated with poorly differentiated tumors in hepatocarcinomas (HCC) (Saito et al., 2001). In an investigation using 24 paired colorectal cancer samples, MBD4 was found to be a significant prognostic factor for patient survival by Kaplan-Meier survival analysis (Xi et al., 2008). In addition, expression levels of MBD4 and MBD3 correlated with the grade of malignancy in human gliomas both in vivo and in vitro (Schlegel et al., 2002). It is not surprising that a defect in this system results contributes to disease pathology, for example by increasing the mutation rate at a second site in the two hits (Knudsen) cancer model.

MBD4 is proposed to be a candidate involved in DNA demethylation activity. This activity might be stimulated in partnership with DNMT3A and DNMT3B, as the enzymatic excision activity of MBD4 is 30–40-fold lower than its T•G mismatch glycosylase activity (Wu and Zhang, 2010). Similarly the only other mammalian thymine glycosylase TDG (Thymine DNA glycosylase) was also reported to interact with DNMT3A and DNMT3B and act as a 5meC glycosylase activity against hemimethylated DNA with the same weak excision activity as MBD4 (Wu and Zhang, 2010) (Figure. 1.6 E). In addition to the possible demethylation activity of MBD4 at specific promoter, zebrafish mbd4 has been proposed to be a candidate protein that is involved in DNA demethylation activity in zebrafish (Rai et al., 2008). Rai and colleagues showed that MBD4 removes G:T mismatch-specific thymines, which results from 5mC deamination, via the coupling of certain enzymes in the cytidine deaminase family (Activation Induced deaminase (AID) and apolipo-protein B RNA-editing catalytic component (Apobec) (Rai et al., 2008b). Interestingly, deamination activity by AID/Apobec may not occur unless MBD4 and/or other possible factor are present and/or activated, and a catalytically inactive hMBD4 derivative (D560A) stabilized the putative G:T intermediate and prevented rapid thymine removal (Rai et al., 2008b). MBD4 was also reported to interact with Xenopus DNMT1 to promote p53 dependent apoptosis (Ruzov et al.,

2009b), and DNMT3b, which is proposed to methylate cytosine and to deaminate 5-meC, relying on an inefficient deaminase activity (Metivier et al., 2008, Kangaspeska et al., 2008). Recent studies in mice have implicated TDG in active DNA demethylation, interacting with the deaminase AID and the damage response protein GADD45a. These findings suggest a two-step mechanism for DNA demethylation in mammals, whereby 5-methylcytosine and 5-hydroxymethylcytosine are first deaminated by AID to thymine and 5-hydroxymethyluracil, respectively, followed by TDG-mediated thymine and 5-hydroxymethyluracil excision repair (Cortellino et al., 2011a, Cortazar et al., 2011). It is possible that MBD4 may make a contribution to this process (Figure. 1.6 F).

Both TDG and MBD4 have been shown to be capable of removing guanine (G):uracil (U) mispair and guanine (G):thymine (T) mispair, products from deamination of the exocyclic amino group in cytosine and 5-methylcytosine. These two coexisting glycosylases seem redundant for a system targeting mismatched thymine and uracil. Their evolutionary timeline is very different (Figure. 1.7). Both glycosylase ancestors of MBD4 and TDG were found in Bacteria. Glycosylase ancestors of MBD4 were also found in Archaea, whereas only a hypothetical protein TDG Archaea ortholog was identified. Both glycosylase ancestors of MBD4 and TDG were identified in plants and fungi. In contrast, the earliest MBD domain ancestors exist in plants but not in fungi. Most importantly, whereas the TDG protein was found in almost all the species throughout evolution, full length MBD4 emerges as a fusion protein only from Chordates onwards, the representative of Invertebrate-Vertebrate transition. This suggests that in a more complex system such as vertebrates, an additional glycosylase like MBD4 is required to maintain genome integrity and MBD4 may possess some special roles such as apoptosis signaling, where fatal mismatches cannot be repaired. Interestingly, according to the evolutionary rate, the proportion of CG sites in vertebrates is far less than that of lower organisms, due to the accumulation of evolutionary repair events of spontaneous deamination. The additional glycosylase system aiming at a lower proportion of CG sites may result in more precise genome surveillance, which is a requirement in higher animals. Alternatively, there might be other functions the additional glycosylase possesses, in order to precisely regulate the system such as the epigenetic transcriptional repression. Recently it has been demonstrated that TDG is essential for early mouse development and the embryonic lethal phenotype of $Tdg^{-/-}$ mice includes mis-expression of developmental genes, suggesting it may have a structural role in maintaining sites of active gene expression (Cortazar et al., 2011, Cortellino et al., 2011a). There was some evidence that TDG might also function to erase aberrant methylation at normally

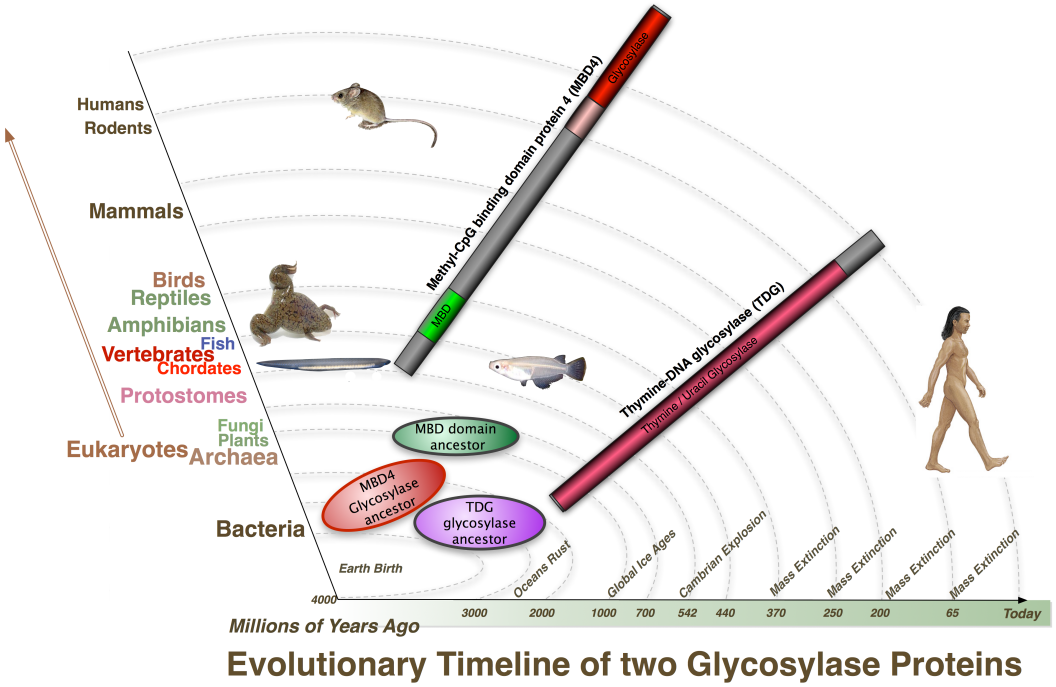


Figure. 1.7. Evolutionary timeline of the two mammalian thymine glycosylase proteins: methyl-CpG binding domain protein 4 and thymine-DNA glycosylase (TDG). The ortholog of proteins including the MBD domain and the glycosylase domain of human MBD4 and full length human TDG were inferred initially by reciprocal best BLASTP searches against the non-redundant protein sequences database, and confirmed by phylogenetic reconstructions. Neighbor-joining (NJ) phylogenetic trees were constructed for domains of MBD4 and full length TDG. Maximum likelihood (ML) analyses were performed using the PHYML module from Geneious pro software, following the JTT model of amino acid substitution. Protein domains were identified using InterProScan and conserved domains servers. Published procedures for this bioinformatic approach were followed (Albalat, 2008). Glycosylase ancestors for MBD4 and TDG can be found in Bacteria. Glycosylase ancestors of MBD4 were also found in Archaea, whereas only a hypothetical TDG Archaea ortholog was identified. Both glycosylase ancestors of MBD4 and TDG were found in Plants and Fungi, in contrast, the earliest MBD domain ancestors were in plants and not in Fungi. TDG protein was found in almost all the species throughout evolution, the full length MBD4 though, emerges as a fusion protein only from Chordates onwards, the representative of Invertebrate-Vertebrate transition.

methylation free CpG island promoters. A bisulfite deep sequencing approach of TDG mutant embryos may address this possibility more fully, with the caveat that the sites may be 5-hydroxymethylated and not 5mC modified. MBD4 was also reported to interact with *Xenopus* DNMT1 in response to DNA damage and DNMT3b (Ruzov et al., 2009b, Boland and Christman, 2008a, Metivier et al., 2008, Kangaspeska et al., 2008) (Figure. 1.6 C). Controversially DNMT3b was proposed to methylate cytosine and to deaminate 5-meC,

relying on an intrinsic inefficient deaminase activity as part of a gene activation program (Ruzov et al., 2009b, Metivier et al., 2008, Kangaspeska et al., 2008). Thymine glycosylases such as TDG and MBD4 may function on deamination of 5-methyl-cytosine by repairing the mismatch.

In sum, MBD4 is a special MeCP protein which possesses a MBD binding domain and a distinct glycosylase domain. It has been associated with a number of pathways including DNA repair, apoptosis, transcriptional repression, and possibly DNA demethylation processes. However, the precise contribution of MBD4 to these processes remains unclear. This led me to undertake the investigation of novel roles of MBD4 as my PhD project.

1.8. Research aims

The aim of the research presented in this thesis was to functionally analyse the methyl-CpG binding domain protein MBD4. Four principal areas of research were investigated:

i) High-throughput screens to identify the interacting protein partners of MBD4 (Chapter 3). The yeast two-hybrid and affinity purification followed by mass spectrometry assays were performed. Additionally, functional annotation of the candidates were used to identify the significant clusters and to select the most robust MBD4 interacting candidates for the further validations.

ii) Validate and character the interactions between MBD4 and the selected protein candidates (Chapter 4). Reciprocal IP and immunocytochemistry was used to analyse the interactions and intracellular localisation of candidates, and the dependence of candidate localisations on MBD4. Additionally, the domains of MBD4 responsible for the interactions were identified.

iii) The biological significance of Mbd4 (Chapter 5). The biological significance of MBD4 was investigated by bioinformatic analysis of the gene expression changes of two independent MEFs cell lines that lack Mbd4 activity in parallel with Tdg, the only other thymine glycosylase in mammals. The potential direct interaction between Mbd4 and estrogen receptor alpha (ERα) was also investigated.

iv) The in vitro binding specificity of MBD domain of MBD4 (Chapter 6). The in vitro EMSA assay was used to determine the evolutionary binding specificity of MBD domain of Xenopus and medaka (Fish) MBD4 on a variety of probes, including unmethylated, methylated, hydroxymethylated, and mismatched DNAs.

Chapter 2: Materials and Methods

2.1. Microbiological techniques

2.1.1. Transformation of bacteria

Chemically competent DH5α *Escherichia coli* (*E. coli*)(Invitrogen) were mixed with an appropriate 10-100ng (1-5ul) of plasmid and incubated on ice for 30min. The transformation mixture was heat shocked in water bath at 42°C for 45sec, immediately placed on ice for 2min, then 500µl S.O.C. media (Invitrogen) of room temperature was added, and the mixture was incubated at 37°C with shaking for 1 hour. Different amount (usually 40ul, 100ul, and the rest all) of transformed cells were aseptically spread onto L-agar plates supplemented with the appropriate antibiotic. The L-agar plates were made in batch by HGU technical services, and the antibiotics concentrations were ampicillin at 50mg/ml; kanamycin at 50mg/ml; chloramphenicol at 13mg/ml. For blue/white selection 80µl X-gal (25mg/ml) was spread onto the agar plates and dried prior to bacterial plating. Streaked plates were inverted and incubated at 37°C overnight.

2.1.2. Growth of bacteria and isolation of plasmid DNA

For the preparation of plasmid DNA, single colonies of transformed bacteria from the selection plate were picked and used to inoculate 3-5ml LB (Luria-Bertani) broth (prepared by HGU technical services) supplemented with the appropriate antibiotic at the concentrations supplied above. The mixed LB broth was incubated at 37°C overnight with shaking at >225rpm. For maxi-preparations, 1.5ml of this culture was used as pre-culture to inoculate 150ml of LB supplemented with appropriate antibiotics and incubated for a further 16hr. 'Nucleospin Plasmid' kit for 3ml cultures (Machery-Nagal) or 'GenElute HP Endotoxin-free Maxiprep Kit' for 150ml cultures (Sigma) were used for plasmid DNA isolation according to manufacturer's instructions.

2.2. DNA preparation and manipulation

2.2.1. Polymerase chain reaction

Specific DNA molecules were amplified from a DNA template by the polymerase chain reaction (PCR) as described by (Mullis et al., 1986).

Routine PCR protocol

50µL reaction	
10xPCR buffer	5µL
Forward Primer	5 µL
Reverse Primer	5 µL
$MgCl_2$	1.5 µL
dNTPs	1 µL
Taq	0.5 µL
Template DNA	1 µL
ddH_2O	X µL

1. 95°C 4’
2. 94°C 30”
3. 57°C 45”
4. 72°C 50”
5. go to 2. 29 times
6. 72°C 5’
7. 4°C

End

2.2.2. Restriction digest of DNA

Restriction digests contained 4U restriction enzyme (Roche)/µg DNA and the appropriate reaction buffer, and were supplemented with 100µg/ml BSA if required. Restriction digests were incubated at the appropriate temperature for the restriction enzyme (usually 37°C) for a minimum of 2 hours. Digested DNA was resolved on a 2% agarose gel, stained using ethidium bromide according to the manufacturer's instructions, then visualised by UV trans-illumination.

2.2.3. Agarose gel electrophoresis and DNA extraction from agarose

The separation of DNA on the basis of molecular weight was achieved by agarose gel electrophoresis using the Sub-cell system (Bio-Rad). 0.8-2% (w/v) agarose gels were used

depending on the molecular weight of the DNA fragment(s) of interest. Agarose gels were prepared with TAE or TBE and contained 0.5µg/ml ethidium bromide, a DNA intercalating agent that fluoresces in ultraviolet (UV) light. DNA samples and100bp DNA ladder (Promega) in Blue/Orange 6X loading dye were loaded into the wells of the gel. Agarose gels were run at constant voltage (50-125Vdepending on the size of the gel) in TAE or TBE electrophoresis buffer, then DNA was visualised under UV light. If a DNA fragment was to be purified from the gel to be cloned, the exposure to UV light was kept to a minimum. DNA fragments required for cloning were resolved by agarose gel electrophoresis and were purified using the NucleoSpin Extract II kits according to the manufacturer's instructions.

2.2.4. Phenol-chloroform extraction and ethanol precipitation

At least 200µl of sample was used for the DNA isolation using an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Sigma), then mixed, allowed to settle followed by centrifuged at 16,000g for 15min at room temperature. The mixture appears aqueous and organic components divided into phases and the top aqueous layer was the DNA-containing aqueous phase, and was carefully taken. An equal volume of chloroform was added followed by pulse vortex and centrifuged at 16,000g for 5 minutes at room temperature. The top aqueous phase was carefully taken to remove residual phenol. 1/10 volume of 3M sodium acetate (Sigma) and 2.5 volumes of ice cold ethanol were added into the aqueous phase to precipitate DNA and incubating for 5min at room temperature or longer at -20°C if required. 1µl of glycoblue (15mg/ml, Ambion) was added when precipitating limiting quantities of DNA according to manufacturer's instructions. Precipitated material was centrifuged at 16,000g for 20min at room temperature and the pellet was washed at least once in 75% ethanol to remove salts. The washed pellet was resuspended in distilled ultra-filtered water (MilliQ water).

2.2.5. DNA quantification and assessment

The concentration of purified nucleic acids was quantified using in a Nanodrop spectrophotometry. The 'Nanodrop ND-1000' was used to measure the A_{260} and A_{260}/A_{230} and A_{260}/A_{280} ratios of DNA samples for quantifying both dsDNA and ssRNA samples. The A_{260} was used to calculate the concentration of the sample (as an OD of 1 at 260nm =

50µg/ml of dsDNA). A_{260}/A_{230} and A_{260}/A_{280} ratios were for estimating the purity of nucleic acids.

2.2.6. pGEM-T Easy cloning

Because *Taq* polymerases often leave a dA 3' overhang on the end of PCR products, PCR products can be cloned into pGEM T Easy vector (Promega) that has 5' dT overhangs. PCR products were ligated with pGEM T Easy vector according to the manufacturer's instructions except 4°C overnight incubation was adopted to get maximum efficiency of ligation.

2.2.7. DNA ligation

DNA ligation reactions for target vectors were performed in a 20µl volume, and contained 5 Weiss units T4 DNA ligase, Rapid ligation buffer (Roche) and 50ng vector DNA. A 3-5:1 molar ratio of insert DNA to vector DNA was present in each reaction. DNA ligation reactions were incubated for 5mins at room temperature, and then 5µl was transformed into a 200µl aliquot of competent bacterial cells.

2.2.8. DNA sequencing

Plasmid DNA to be sequenced was purified using a NucleoSpin Extract II kit according to the manufacturer's instructions. DNA sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing kit (ABI). DNA sequencing reactions contained 2µl BigDye Terminator v3.1, 3.2pmol primer and 50-300ng DNA template in a 10µl volume, and were incubated under the following conditions: initial DNA denaturation at 96°C for 1min, then 25 cycles of DNA denaturation at 96°C for 10s,primer annealing at 50°C for 5s, and DNA polymerase extension at 60°C for 4min.Sequencing reactions were cleaned up and were run on an ABI capillary sequencer by the MRC Human Genetics Unit Sequencing Service.

2.2.9 in vitro methylation

Plasmids were mixed with the appropriate bacterial methylase and buffer (NEB) supplemented with 160µM S-Adenosyl methionine (SAM) or 32µM SAM for HhaI methylase at 37°C for 2 hrs. Fresh SAM was often added after 2hrs and the sample incubated for a further 1hr at 37°C. The reaction was stopped by adding the breakdown product of SAM, S-adenosyl-L-homocysteine (SAH), to inhibit the methyltransferase reaction. Phenol/chloroform extraction was used to purify the products. For confirmation of the in vitro methylation, the purified products were digested with the methyl-sensitive enzyme HpaII and its methyl- insensitive isoschizomer MspI.

2.2.10. Site-directed mutagenesis

The QuikChange® II XL Site-Directed Mutagenesis Kit (Stratagene) was used according to the manufacturer's instructions to introduce base changes/insertions in the sequence of plasmid DNA templates. Mutagenic oligonucleotide primers were designed by Stratagene online program available at http://www.stratagene.com/qcprimerdesin. Reactions were set up at 50µl volume and either 10ng or 50ng of dsDNA template were used. The reactions were then cycled through the following temperature cycles in a thermal cycler:

Segment	Cycles	Temperature	Time
1	1	95°C	1 minute
2	18	95°C	50 seconds
		60°C	50 seconds
		68°C	1 minute/kb of plasmid length + 2 minutes
3	1	68°C	7 minutes
4		4°C	

On completion of the thermal cycling program, 1µl of Dpn I restriction enzyme (10 U/µl) was added directly to each amplification reaction, then reactions were gently mixed and incubated at 37°C for 1 hour. Either 2 or 10µl of the digested mixture was used to transform XL10-Gold ultracompetent cells provided by the kit. Transformed clones were selected on agar plates containing appropriate antibiotics, and then plasmid DNA was prepared and checked by sequencing for the required nucleotide sequence alteration.

2.3. RNA preparation and analysis

2.3.1. Isolation of RNA from mammalian cells

'Trizol' reagent (Invitrogen) was used for RNA isolation from mammalian cells and tissues according to manufacturer's instructions. This is usually suitable for the occasion where cell numbers were not limiting (>1 x 10^4). Briefly, Trizol was added to adherent cells directly onto the culture surface (after removing media) to minimise chance of RNA degradation. Usually per 25cm^2 of culture surface for mouse embryonic fibroblasts at >70% confluence, 1ml of Trizol was added. The samples were then incubated for 2 minutes at room temperature to disrupt nucleoprotein complexes, and then 1/10 volume of bromochloropropane (BCP) was added. After mixed vigorously by hand, the mixture was allowed to briefly settle down, and immediately centrifuged at 12,000g for 15min at 4°C. The centrifuged sample contains separated organic, inter- and aqueous phases. The top RNA-containing aqueous phase was carefully transferred to a fresh RNase free eppendorf and precipitated with an equal volume of isopropanol (if RNA quantity is limited, 1ul glycoblue (Ambion) was added) for 10 minutes at room temperature. The sample was pelleted by centrifugation at 12,000g for 10 mins at 4°C, and then washed in 75% ethanol. The washed pellet was resuspended in RNase-free water (DEPC treated water - Sigma), and a portion of RNA was taken for determining the concentration by spectrophotometry using Nanodrop ND-1000 on RNA settings, and the RNA integrity by ethidium bromide agarose gel electrophoresis for the abundant ribosomal RNA fragments or by Bioanalyser RNA1000 chip (Agilent) for accurate integrity accessment (Bioanalyser analysis performed by HGU technical services). The rest portion of resuspended sample was stored at –80°C for long-term storage. Generally, the utmost care is needed for handling RNA. Clean gloves, pipettes and bench surfaces were required. Where possible, the working spaces were rinsed with RNase Zap (Ambion), and RNase-free filtered pipette tips, eppendorf tubes and water were used throughout.

2.3.2. Preparation of cDNA

Where possible, all RNA samples were DNase treated with Turbo DNA-free DNase (Ambion) according to the manufacturers instructions for the generation of complimentary DNA (cDNA). When RNA quantities were limiting, the DNase step was not performed occasionally, but very top RNA-containing aqueous phase was carefully transferred to ensure there is no DNA contamination, and in subsequent reactions negative controls are

required such as -RT (no reverse transcriptase). Where possible, 2-5ug RNA was used in each DNase reaction as a template for cDNA (complimentary DNA) synthesis. The reverse transcription reaction was primed by random heximers (Promega) using SuperscriptII or SuperscriptIII (Invitrogen) according to manufacturer's instructions, and RNase inhibitor (RNaseIn - Promega) was supplemented to the reaction to ensure no degradation of the RNA template in the early stages of the reaction. Negative control of cDNA reactions without the addition of the reverse transcriptase was performed to ensure there are no DNA and aerosol contamination, and in downstream experiments non-intron-spanning primers were used. cDNA was stored at -20°C, and -70°C for long-term storage.

2.3.3. RT-PCR

RT-PCR using cDNA as a template was performed to semi-quantitatively assess the specific genes in steady-state cellular RNA transcription level. 'Primer3' program (http://frodo.wi.mit.edu/primer3/) was used for primer desig (Table 2.1). Brifly, the default settings were adopted, usually a T_a (annealing temperature) of 55°C, a product size of <400bp and skipping a large intron where possible were used to avoid genomic DNA contamination. RT-PCR was performed with 1x PCR buffer (Invitrogen), 2mM $MgCl_2$, dNTPs (0.5mM each), 150nM each primer, 0.6u *Taq* (*Thermus aquaticus*) DNA polymerase (Invitrogen) and an appropriate volume of cDNA. Reactions were incubated in a thermocycler using the following program:

Segment	Cycles	Temperature	Time
1	1	95°C	2 minute
2	18-35	94°C	15 seconds
		56°C	20 seconds
		72°C	25 seconds
3	1	72°C	10 minutes
4		4°C	

PCR products were run by agarose gel electrophoresis and visualised using ethidium bromide staining and UV trans-illumination.

2.4. Protein preparation and analysis

2.4.1. Cell extracts

Whole cell protein extracts were prepared by adding an appropriate volume of Lammelli buffer in PBS washed cells (60mM Tris pH6.8; 10% glycerol; 2% SDS w/v; 0.01% bromphenol blue w/v; 10mM DTT). The mixture was sonicated briefly using an 'Soniprep 150' probe sonicator (MSE)(3 pulses of 30s at 10µm amplitude), centrifuged to remove debris (5min at 20,000g at room temperature) to sheer chromatin and boiled for 5 minutes at 95°C to denature proteins. The supernatant was taken and glycerol added to a 10% final concentration, before storage at –20°C.

2.4.2. SDS-PAGE

The separation of proteins on the basis of molecular weight was achieved by denaturing SDS polyacrylamide gel electrophoresis (SDS PAGE) according to the method of Laemmli (1970), using the vertical electrophoresis Mini-PROTEAN system (Bio-Rad). The polyacrylamide separating gels were between 8% and 15% (w/v) depending on the molecular weight of the protein(s) of interest. 5% (w/v) polyacrylamide stacking gels were used. Protein samples were mixed with 2X SDS PAGE loading buffer and Pre-stained protein marker (Protein Loading Buffer Pack, Fermentas), were incubated at 95°C for 4 min and then put on ice, before being loaded into the wells of the stacking gel. SDS PAGE gels were run at constant current (200V) in Tris-glycine electrophoresis buffer. SDS PAGE gels were stained by incubating in GelCode Blue Stain Reagent (Thermo) with shaking for 1 hour in room temperature or in cold room overnight. Gels were washed in diswater until protein bands were clearly visible. The stained gels were dried if a permanent record was required.

2.4.3. Western blotting

Proteins were transferred from SDS PAGE gels onto Hybond-ECL Nitrocellulose membrane (Amersham) using a GENIE® Electrophoretic Transfer Blotter (ideascientific) and fresh Methanol-Tris-Glycine transfer buffer. Blotting was performed at 15V 2.7A constant current (TELEX AC.DC power supply) for 45 mins depending on the surface area of the gels. Non-specific binding sites on the western blot were blocked by incubation in 5%(w/v) Marvel milk powder in 0.5% TBST at 4°C o/n, or for at least 1h on a shaking platform at r/t. All subsequent steps were also performed on a shaking platform at r/t. The blot was washed 2

times 15 mins 0.5% PBST and then incubated with a primary antibody in 0.5% PBST for 2h or in cold room overnight. After 3 times 5 mins washes in PBST, the blot was then incubated with either an Alkaline Phosphatase Conjugated or a horseradish peroxidase (HRP)-conjugated secondary antibody in 5% (w/v) milk in 0.5% PBST for 1h, then washed 2 times 15 mins 0.5% PBST and 2 times 15 mins PBS washes. The blots were incubated either by Immun-StarTM AP Substrate (Bio-Rad) for 5 mins or Equal volumes of ECL solutions 1 and 2 were mixed, and the composite ECL solution was incubated with the blot for 1min with gentle shaking. The blot was placed in a developing cassette, covered with Saran wrap and exposed to Hyperfilm ECL (GE Healthcare) for 10 mins (AP) or 30s(ECL), then for shorter or longer periods of time as required. Films were developed by a SRX-101A tabletop film processor (Konica Minolta). Primary and secondary AP-, HRP-conjugated antibody dilutions were based on recommendations from the antibody supplier(s) and on empirical observations.

2.4.4. Purification of recombinant protein

Ni-NTA agarose purification of His-tagged proteins produced in E.coli were purified by nickel agarose chromatography. Ni-NTA beads (Invitrogen) were prepared according to manufacture's instructions, and then mixed with the cleared supernatant from the lysed cells. Bacterial cell lysates were incubated with the nickel agarose beads on a roller at 4°C for 1 hour to allow binding of His-tagged proteins to the resin. After binding, the lysates with beads were transferred to a reusable Polyprep Chromatography Column and the lysates were loaded at least 3 times. The beads were then washed once in 10 times beads volume Washing Buffer A, PH8.0. Proteins were eluted from the column by appropriate volumes of Elution Buffer A for 6-8 fractions.

2.5. Mammalian cell culture

2.5.1. Culture of mouse embryonic fibroblasts

Prior to defrosting mammalian cells, the appropriate cell culture medium was prewarmed to 37°C. The stock of frozen cells was quickly thawed in a 37°C water bath. 5ml medium was added drop-wise to the cells, and the cell suspension was incubated at r/t for 2min. A further 10ml medium was added drop-wise. The cells were then pelleted by centrifugation at 1,000rpm for 4min. A single cell suspension was prepared in fresh medium and transferred to a T75 cell culture flask (12ml medium/T75 flask). The volume of medium and the diameter of the flask used were determined by the amount of cells in the stock. Flasks were incubated at 37°C in 5%CO2. When the cells were confluent, the plate was washed with PBS before being incubated with trypsin-EDTA (2ml/T75 Flask) for 2 mins. Trypsin is aserine protease that can mediate the detachment of cells from the cell culture plate surface. After cell dissociation from the surface, fresh medium (10ml/T75 plate) was added, and a single cell suspension was prepared. The cells were divided by either 1/6 or 1/8 and were resuspended in fresh medium. The required volume of the cell suspension was transferred into a new cell culture flask with an appropriate volume of additional medium. As commonly used mammalian adherent cell lines divide approximately once every 24h, and cells were routinely passaged to 12.5-25% confluency, passaging was carried out once every 2-4d.When preparing a stock of frozen cells, cells were trypsinised and spun down as described above. The cell pellet was then resuspended in cell culture medium containing 10% (v/v) DMSO (cells from a T75 flask were resuspended in3ml total volume). The presence of DMSO increases the percentage of cells that survive freezing and thawing by minimising the formation of ice crystals within the cells. The cell stocks were stored at −20°C for 1h, then at −70°C o/n, before being transferred to a liquid N2 cryostore for indefinite storage.

2.5.2 DNA transfection

HEK293T, CMT93, or MEFs Cells were seeded at appropriate number at least 24hrs prior to transfection in culture media lacking antibiotics and grown to a confluence up to 70-80%. Lipofectamine 2000 (Invitrogen) was used in accordance with manufacturer's instructions.

Briefly, plasmid DNA and lipofectamine were added in separated appropriate volume of Opti-mem (Gibco) for 5 minutes and then combined together with gentle mixing. The combined solution containing DNA and lipofectamine in Opti-mem was incubated for a further 25 minutes before being added directly to cells (up to 30 minutes). Appropriate amount of Lipofectamine 2000 was used according to manufacturer's instructions.

2.6. Experimental procedures

2.6.1 Yeast Two Hybrid Mating Procedure

Reference: Bendixen et al. (1994) **Nucl. Acids Res.** 22 (**9**):1778-1779 See also in Figure. 3.2

1. Start an overnight culture of 30 ml of SD –Trp from a single yeast colony containing GAL4 BD-Bait construct (pGBKT7). Use a colony less than 3 weeks old for best results. Shake at 30°C.
2. The next day, ~24 hours after inoculation, use the 30 ml culture to inoculate a 2 X 300 ml culture of SD –Trp (i.e. 270 ml of YNB 30 mls of SD –Trp) with 150 ul and 300 ul, respectively. Shake overnight at 30°C.
3. The next day in the morning measure the OD600 and choose the culture closest to 0.7 to 0.9 to continue with.
4. Take a frozen aliquote of the library cDNA and resuspend directly with 1 ml of YPD (pipetting up and down) and add to 50 ml of YPD. Incubate at 30°C for 30 minutes.
5. Mix library (50ml) with 300 ml culture and rinse the library flask several times with liquid from the 300 ml culture. Collect cells on filters (0.22 mm GS (millipore)) in screwtop bottle top by syringing through the cells in 20 ml aliquotes. Follow with a syringe of 10 ml YPD and a syringe to push through air until there is no more liquid.
6. Place the filter discs, cell side up, on YPD plates (place 4 discs on each 15cm YPD (+Ade) plate) and incubate at 30 c for 5 hours.
7. Resuspend cells on filters by placing them in a sterile petri and rincing up and down with 2 X 750-1000 ul of YNB (depends on number of plates). Pool all the washes in one 50 ml falcon.
8. Plate 1-1.2 mls of resuspended cells per 24cm X 24 cm library plate which is SD-Trp -Leu and –His (2 mM AT-3) use glass beads to spread and incubate at 30 c for 5 days. Colonies should appear by then.

9. As a control dilute 100 ul of resuspended cells in 1ml of YNB for a 1/10 dilution and make 1/200 and 1/4000 from that dilution using YNB. Then plate 100 ul of 1/10 and 1/200 dilution on SD-Trp and SD-Leu plates and 1/200 and 1/4000 on SD-Trp/-Leu (gives idea of efficiency of the mating). In 2 days you should have colonies.
10. After 5 –6 days pick any yeast colonies that grow and replate on SD-Ade, SD-His and SD-Ade/His plates containing 2 mM AT. Good colonies should grow on all three. Interaction can also be tested by increasing concentration of AT to 5 mM, which select for only strong interactors. Do this by picking a colony into 50-100 ul of dH2O and laying a drop of about 3-5 uls on each plate (use matching grids).

Add 2 ml of AT per litre of agar and 100 mls of SD to give SD-Ade/His then add either 1 ml/100ml of His (2mg/ml) or Ade (10 mg/ml) for –Ade or –His, respectively (note: SD is –Trp/Leu/His/Ade to begin with).

2.6.2 IP and MS of MBD4-Containing Molecular Complexes

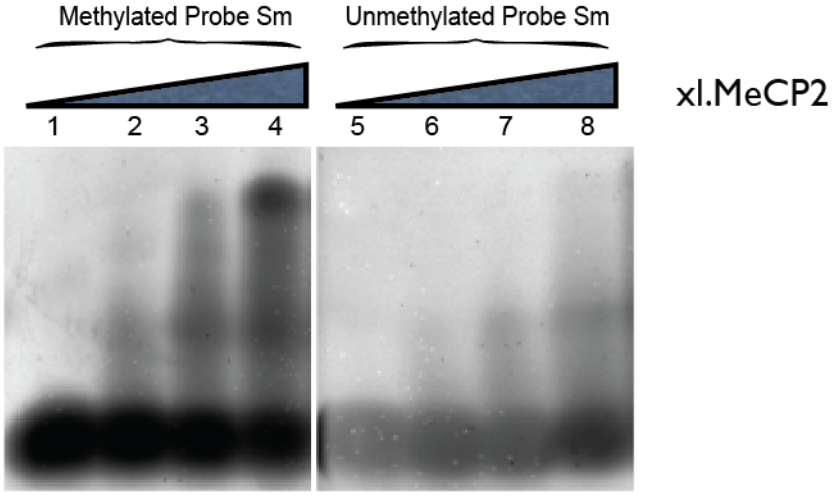
HEK293T cells were seeded to fifty 10cm tissue-culture plates. When the cells grow up to 70% confluence, the cells were transfected with FLAG tagged expression vector. About 48 hr posttransfection, cells were lysed with a hypotonic lysis buffer (0.05% NP-40, 10 mM HEPES, 1.5 mM MgCl2, 10 mM KCl, 5 mM EDTA, and complete protease inhibitor cocktail (Roche), pH 7.4, $30x10^6$ cells/ml), and were incubated on ice for 15 min. Cytosolic fraction was discarded and the separated cell nuclei were incubated on ice for 1 hr in a nuclear extract buffer (20 mM HEPES, 300 mM NaCl, 20 mM KCl, EDTA-free complete protease inhibitor cocktail (Roche) and MNase (Nuclease S7; Roche), pH 7.4, $30x10^6$ cells/ml). a final concentration of 5mM EDTA was used to stop the chromatin digestion, and the sample was centrifuged at 20,000g for 30 minutes twice to get post nuclear supernatants. 50ul Sepharose beads covalently conjugated to Flag- specific mAb (Sigma) or were added to Samples and were incubated for 2 hours with rotation at 4°C. Beads were washed (3x) with ice-cold PBS containing 0.05% NP40, and once with pure ice-cold PBS. Bound proteins were eluted by boiling for 5 min in sample buffer, and the eluted samples were loaded on a large 10% SDS-PAGE (BioRad) and separated. Gels were stained with Colloidal Blue (NuPAGE, Invitrogen) Reference: (Abramson et al., 2010). Several chunks of bands of diverse molecular weights were excised from the experimental and control lane of a corresponding molecular weights. The gel chunks were sent to St. Andrews Mass Spec services for analysis.

2.6.3 Electrophoretic mobility shift assay (EMSA)

All oligonucleotides were synthesized by Sigma- Aldrich and purified by desalting. Double stranded probes for EMSA assays were generated by annealing a reversed single stranded oligonucleotide to the appropriate forward single stranded oligonucleotide. The resulting probes were purified by 8% none denature polyacrylamide gel, and eluted by Maxam and Gilbert elution buffer. The purified probes were then methylated with M.Sss1 methyltransferase to obtain methylated probes with a single methylated CpG (m+). The double stranded probes were purified by Phenol-chloroform extraction and end labelled with T4 polynucleotide kinase (Promega) in the presence of gamma 32P labelled ATP. EMSA reactions were assembled in 20 µl reaction volumes at room temperature by pre-incubating 500 ng poly(dG-dC)•poly (dG-dC) (Sigma) or Ecoli (Sigma) competitor DNA and recombinant protein in binding buffer (8 % ficoll, 20 mM HEPES pH 7.9, 150 mM KCl, 1mM EDTA, and 0.5mM DTT) for 10 min. In each EMSA recombinant proteins were usually used from 100nM to 1000nM (from 100nM at a 2X increasing progressively manner, for example, 100nM, 200nM, 400nM, etc) (Figure. 2.1). Labelled probe DNA containing bromophenol blue as a migration marker was added and the reaction which was allowed to continue for a further 25 minutes at room temperature. The binding reaction was loaded on to a 6 % or 8 % non-denaturing polyacrylamide that had been pre run for 1 hour at 250 V in 0.5X TBE at 4 °C. The samples were electrophoresed at 250 V for 2hours, until the bromophenol blue dye front had migrated 10 cm into the gel. The gel was transferred to 3 mm Whatman paper and covered by Saran, and then dried at 80 °C for 2 hour. Dried gels were exposed to a Phosphor Imager screen overnight and developed the next day.

A

Without Competitor



B

With Competitor

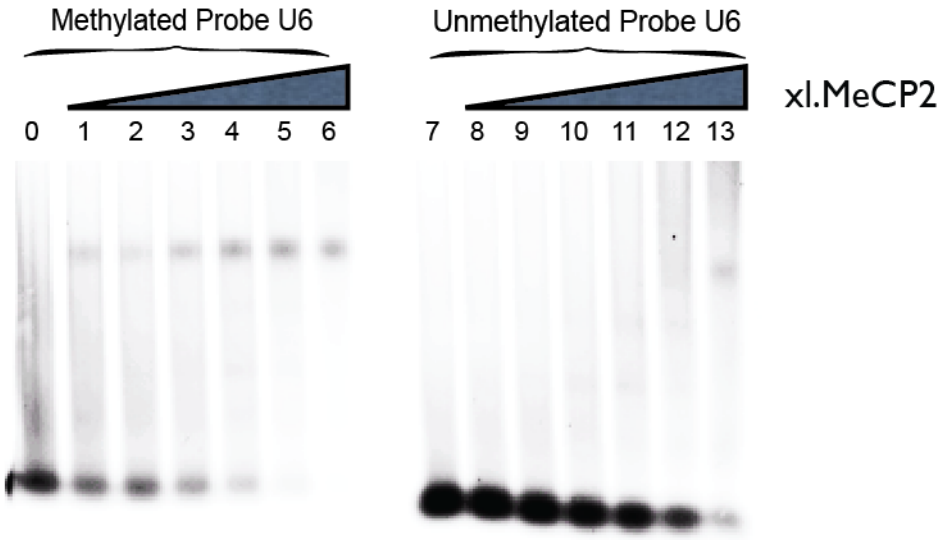


Figure. 2.1. EMSA assay optimisation with and without E.coli competitor (1µg), demonstrating the robust preferential binding specificity of xl.MeCP2 to the methylated probes containing CpG dinucleotides than the responsive unmethylated ones. A. Without competitor, binding of the duplex probes containing 5-methylcytosine (5mC)(left) or the responsive unmethylated cytosines (right) in the context of CpG to varying concentrations of the methyl-binding domain (MBD) of xl (*Xenopus laevis*) MeCP2 from 1&5. Free probe; 2&6. 100 nM; 3&7. 200nM; 4&8. 400nM; sequence of Sm duplex is listed, M indicates the methylated cytosine base. B. With 1µg E.coli competitor, increasing amount of xl.MeCP2 from 50nM (Lane 1&8), 100nM (Lane 2&9), etc. to 1600nM (lane 6&13, excess amount) were co-shifted with methylated U6 probe (left), and unmethylated U6 probe (right, a probe without AT run in comparison to S1 and S2 probes, see chapter 6). In general, from both essays with and without competitor, preferential binding specificity of xl.MeCP2 to methylated probes containing CpG dinucleotides can be clearly observed. The assays with competitors are cleaner and clear binds without smear migrant are usually observed. It also excludes the non-specific binding possibility in the assays without competitors.

2.6.4. Quantitative polymerase chain reaction (qPCR)

Quantitative real-time polymerase chain reaction (qPCR) was used for quantification of cDNA. Reactions were performed in duplicate in white 96-well plates (Eurogentec) in LightCycler (Roche). 10µl of '2x Brilliant II qPCR master mix' (Agilent) was used and supplemented with 150nM each primer, DNA or cDNA template in MilliQ dH_20 for a total 20ul. The program used for the reactions was: 95°C for 5min, then cycles of 94°C for 15sec, 55°C for 20sec and 72°C for 20sec, followed by a melt curve ramping from 60-95°C. Roche LightCycler 480 qPCR machine was used and fluorescence signal of the SYBR green dye was measured after each cycle of PCR. Primers was designed with an annealing temperature of 55°C and a product size of <150bp, and to span a large intron where possible to avoid genomic DNA contamination using the 'Primer3' program (http://frodo.wi.mit.edu/primer3/) (Table 2.1). A standard curve was generated by 10-fold serial dilution of a cDNA sample, and housekeeping genes, such as *actb* and *gapdh*, was used to normalise for loading and overall efficiency of the cDNA synthesis reaction. Technical or biological replicates were performed all the time, the qPCR error (error between the two qPCR replicates) was ignored and the standard error was calculated between the technical or biological replicates. Error bars denote the standard error of the mean (S.E.M) for each sample.

2.6.5. Immunofluorescence of fixed cells

Mouse colon cancer cells CMT93 were seeded on gelatinised glass coverslips at a density of $2*10^5$ cells per well of a 6-well plate and cultured for 24hr. The co-transfection with vectors expressing epitope- tagged (GFP or FLAG) full length Mbd4 with epitope-tagged (GFP, Cherry, or FLAG) interacting candidate proteins. After washed twice in PBS the cells were then fixed in 4% formaldehyde (v/v in PBS)(Sigma) for 10min at room temperature with gentle rocking. Cells were rinsed 3 times in PBS to remove formaldehyde and then 0.1% TritonX-100 (v/v in PBS)(Sigma) was added for 5min at room temperature for cell permealisation followed by 2 rinses in PBS. 100µl of block solution (PBS; 0.5% BSA w/v; 0.1% Tween20 v/v (Sigma); 10% goat serum v/v) was added, and the coverslips were kept in a humidified box and and incubated for 30min at room temperature. The primary and secondary antibodies were diluted in Goat serum as the secondary antibodies used were raised in goat, for minimising background signal produced by secondary antibody binding. Primary antibodies (Antibody table 2.2) were diluted 1:100-1:1000 in block solution and 100µl was added to coverslips. Clean parafilm tailored to thin plastic strips were used to cover the coverslips to prevent them drying. The coverslips were kept in humid box and incubated overnight at 4°C. The second day, coverslips were washed 3 times for 5min in PBS with 0.1% Tween20 v/v (Sigma). Secondary antibodies (Goat α-rabbit IgG-Alexafluor Red conjugate (Invitrogen); Goat α-mouse IgG-Alexafluor Green conjugate (Invitrogen)) were diluted 1:1000 in block solution added to coverslips in the dark for 1hr at room temperature. Coverslips were then washed 3 times for 5min in PBS with 0.1% Tween20 v/v (Sigma), and care need to be taken to minimise the light exposure. Coverslips were mounted in 'Vectashield' mounting media (Vector Laboratories) (250ng/ml DAPI pre-mixed) and viewed using 2 or 3 colour fluorescence microscopy using defined exposure settings. Fluorescence images were taken using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics), Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives and Chroma #83000 triple band pass filter set (Chroma Technology), and 'IP Lab' software was used for image analysis.

2.6.6. Protein extraction from E. coli

Standard expression conditions for protein extraction and elution are as follows.

1. Grow a 5-ml overnight culture at 37°C in BL21-CodonPlus®(DE3)-RIPL cells in Luria-Bertani (LB) broth supplemented with 50µg/ml ampicillin.

2. Subculture this 1:100 into prewarmed 400 ml LB/ampicillin in a baffled 2-liter flask and shake at 37° at 250 rpm for 3 h (to mid/late exponential phase, OD600=0.6 to 0.8). Lower temperatures can also be used to promote protein solubility; we have used 18, 25, and 30° depending on the individual protein.

3. Add isopropyl-β-d-thiogalactoside (IPTG) to 0.4mM and continue shaking for 4–6 h. The IPTG concentration and length of induction can also be varied to optimize the yield of the soluble protein.

4. Harvest cells and resuspend in either 20 ml BugBuster® Protein Extraction Reagent (Novagen) with 1μl/ml Benzonase Nuclease (Novagen), or buffer A [50 mM Na2HPO4/NaH2PO4, pH 7.4, 300 mM NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol (2-ME), 30% glycerol] containing 1 mM 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF). The cells were then lysed by sonication (3× 1-min bursts on ice with a microtip) or without sonication if use BugBuster reagent. Recover the soluble fraction after centrifugation for 30 min at 15,000g at 4°C.

5. Purify the his-tagged protein by a standard protocol of Nickel agarose purification. Proteins are generally eluted with either a Tris- or a phosphate-based buffer containing 200 mM imidazole, 150–300 mM NaCl, 1 mM 2-ME, and glycerol if required.

6. After SDS–PAGE analysis to judge yield and purity, dialyze the eluted protein against dialysis buffer (50 mM Tris- or phosphate-based buffer, pH 8.0, containing 150–300 mM NaCl and 1 mM 2-ME, and glycerol as necessary) to remove imidazole.

2.6.7 Immunohistochemistry on human primary cancer tissues

Tissue microarray (TMA) slides were prepared at 4μm sections, and were deparaffinised in xylene and rehydrated in 99%, 80% and 50% ethanol. For antigen retrieval, 0.15mM Sodium Citrate, pH6.0 buffer was used. Microwave pressure cooker was used for heat induced antigen retrieval for 5min. The sections were cooled down for 10 minutes, followed by rinse in 0.05% phosphate buffered saline (PBS, OXOID, #BR0014G)-Tween (Tween20, BIO-RAD, #170-6531). The sections were blocking in 3% Hydrogen peroxide for 10 minutes and serum free protein block (Dako, #X0909) for 10min. The antibody was diluted in antibody diluent (Dako, #S0809) with 1:100-1:4000 dilution to determine the optimum concentration. The sections were incubated with primary antibodies for either 1 h at room temperature or overnight at 4°C. Dako REAL EnvisonTM Detection system (Dako, #K5007) was used for staining, which consists of Envision/HRP solution, diaminobenzidine (DAB chromogen) and substrate buffer containing hydrogen peroxide. The EnvisionTM/HRP was added onto the

sections for 30 min at room temperature, followed by DAB diluted in substrate buffer at a 1 in 50 dilution for 10 min. The sections were then counterstained in Haematoxylin and Scott's tap substitute. Negative control that has no primary antibody was performed

2.6.8 Immunofluorescence on human primary cancer tissues

The procedures before the incubation of primary antibody were exactly same as described above. Slides were then incubated with primary antibodies along with pan-cytokeratin (Mouse anti-pan cytokeratin or Rabbit anti-pan cytokeratin) diluted in antibody diluent (Dako, #S0809) for 1 h at room temperature or overnight at 4°C. After rinsing in 0.05% PBST, the secondary antibodies (Alexa 555–conjugated goat anti-mouse or goat anti-rabbit antibody diluted 1:100 which have conjugated with horseradish peroxidise HRP, Dako, #K4001, #K4003) were added and incubated for 1 h at room temperature. The sections were then incubated with TSA conjugated with Cyanine 5 (Perkin Elmer, #NEL745) for 10 min to visualize the target protein. The sections were mounted with mounting medium (Invitrogen, #P36931) containing (DAPI, Invitrogen, #P36931). The slides were dried overnight, and then were sealed with nail polish. For long-term storage, slides were kept in dark box at 4°C.

2.6.9 Expression microarray

Total RNA was extracted from each cell line at four time points in two weeks for biological replicates. The Bioanalyzer 2100 (Agilent) quantitative on-chip electrophoresis was used for confirmation of the integrity of RNA samples. For mouse Ref-6 v2 Expression BeadChips (Illumina), biotinylated cRNA for hybridisation was generated using the TotalPrep RNA Amplification kit (Illumina) according to the manufactures instructions. 500ng total RNA was used to synthesise double stranded DNA (dsDNA). Biotin labelled cRNA was generated using purified dsDNA as a template from 6 hours at 37°C in vitro transcription. The biotinylated cRNA was purified and before apply them onto illunina arrays; another confirmation of integrity using the Bioanalyzer 2100 was performed. The application of illumina arrays was performed by the Wellcome Trust Clinical Research Facility, Edinburgh.

2.7. Bioinformatic analyses and data processing

In general, two methods of the Illumina Beadstudio gene expression module v3.4 and Matlab bioinformatics toolbox were used for the bioinformatics analysis and data processing in this section.

2.7.1 Microarray analysis

Beadstudio Gene Expression Module v3.4 (Illumina) was used to analysis the raw data generated from microarray. Average normalization and/or quantile normalization were adopted with background removal (a set of quantile normalization plottings were in Figure. 2.2 & 2.3 & 2.4 to show the data plottings before and after normalizations). Bar and line charts were drawn in Matlab and Microsoft Excel with error bars representing standard error of the mean. MAXY plots were drawn in Matlab. Boxplots were drawn in Matlab using the default settings for whisker length and box height. T-test error model with multiple testing corrections using Benjamini and Hochberg false discovery rate was adopted for analysis of differential expression. In this occasion, two-tailed Student's T-test was performed to test if independent samples are equally large for small numbers of samples or where normality could be assumed. For the non-transformed MBD4 knockout cell line, only two pairs of samples were microarrayed because of limited cost, and Mann-Whitney-U (MWU) test was performed. Mann-Whitney-U (MWU) test (also known as Wilcoxon-Rank-Sum test) is a non-parametric test that was used to test if two independent samples are equally large. MWU test was used where sample size was sufficiently large and a normal distribution of samples could not be assumed.

For very high-stringent criterion, a primary filter of >5-fold expression change threshold was set to determine genes most significantly differentially expressed between samples. The detection p- value in the raw data is calculated based on their relative signals to negative controls and the genes that have detection p-values of $p < 0.05$ can be considered statistically true detected. The expression fold-changes of all the genes (in gene ranking table) and probes (in probe ranking table) whose detection p-values <0.05 were ranked. The genes that expression fold change larger than 5 fold in both MBD4 knockout cell lines were considered as most robust gene candidates. In addition, a 2-fold change stringent was also used, and the gene candidate tables for each cell line were listed in supplementary table 5.1 &5.2.

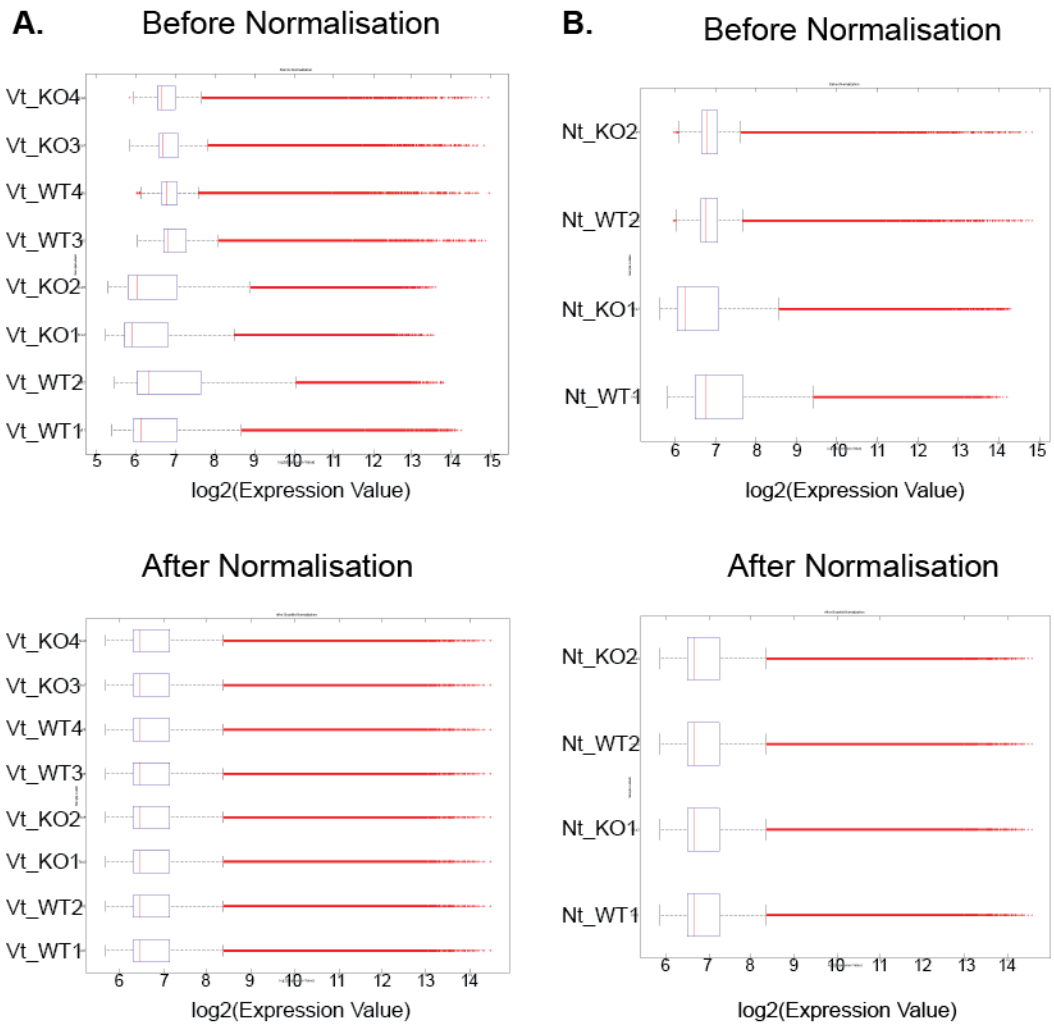


Figure 2.2. Data before and after Quantile normalisation. Boxplots of the log2 transformed signal intensities are shown for each replicate (Vt_WT: virus transformed cell line, wildtype; Vt_KO: virus transformed cell line, Mbd4 knockout; Nt_WT: non-transformed cell line, wildtype; Nt_KO: non-transformed cell line, Mbd4 knockout). The Quantile normalisations rank the data for each array, take the average value at each rank and then for each array replace the expression values with their rank average value. The upshot of this is that the data in each array then has exactly the same distribution. Note there are differences in intensities between samples on the same chip and samples on different chips. The first BeadChip, containing samples Vt_WT&Vt_KO1&2, and Nt_WT&Nt_KO1, seems to be slightly more variable than the second BeadChip containing samples Vt_WT&Vt_KO3&4, and Nt_WT&Nt_KO2. After normalization, the average and length of each sample are aligned.

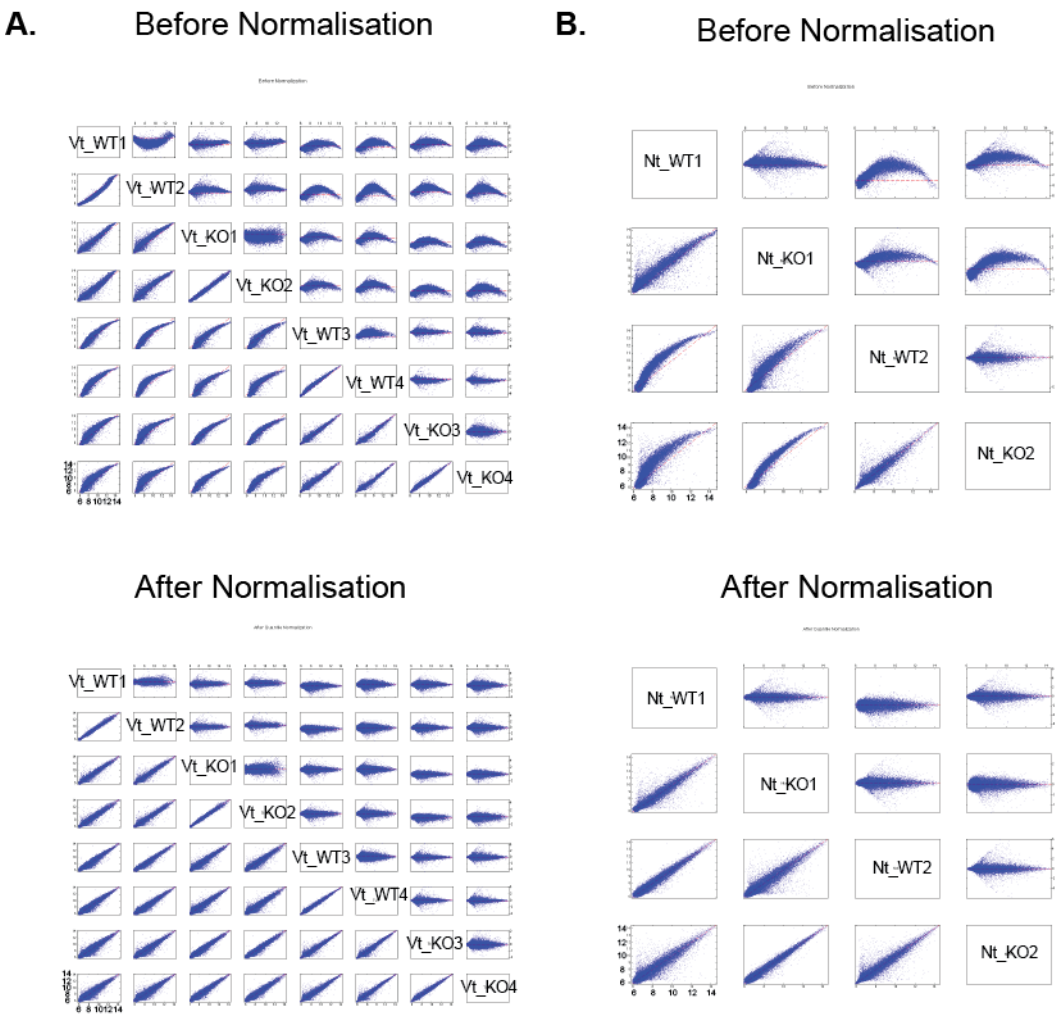


Figure. 2.3. MA and XY plots showing pairwise comparison of the arrays on a BeadChip before and after Quantile normalization. The MA plots for all pairwise comparisons are in the upper-right corner. In the lower-left corner are the XY plots of the comparisons. On an MA plot, the average of the expression levels of two arrays 'A' are plotted on the x axis, and the difference in the measurement 'M' on the y axis. An XY plot is a scatter plot of one array against another. The expression boxplots and MAXY plots reveal that, there are differences in expression levels within chips and between chips; hence, the data requires normalization, and was applied with the Quantile normolisation function.

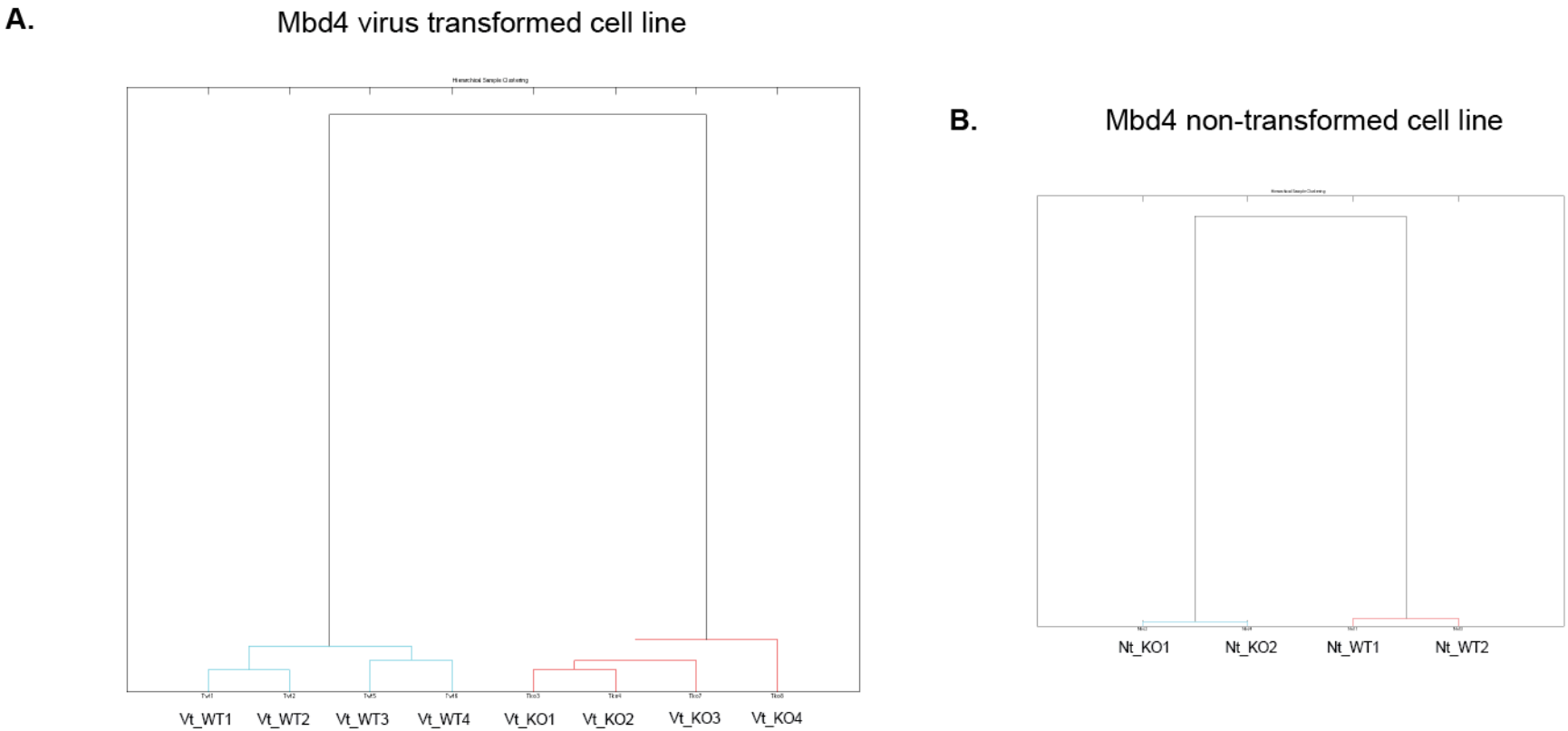


Figure. 2.4. Hierarchical clustering of differentially expressed genes for all arrays. As described in Figure. 2.1. The first BeadChip contains samples Vt_WT&Vt_KO1&2, and Nt_WT&Nt_KO1, while the second BeadChip containing samples Vt_WT&Vt_KO3&4, and Nt_WT&Nt_KO2. Unsupervised hierarchical clustering of the significant gene profiles were performed for differentially expressed genes of the arrays using correlation as the distance metric to cluster the samples. All the wildtype cells were clustered from knockout cells regardless of the different BeadChip where they were arrayed, showing the robustness of the assays and bioinformatic analysis.

2.7.2 Gene ontology (GO) and IPA pathway analysis

The program Database for Annotation, Visualization, and Integrated Discovery (DAVID) 2007 (Dennis et al. 2003) was used for identification of significant gene clusters. Default parameters were adopted, and the background was normalised to Illumina BeadChip v2 Ref-6. Ingenuity Pathway Analysis (IPA) system (Ingenuity Systems, Mountain View, CA, USA, www.ingenuity.com) was used for the pathway analysis and visualisation. The datasets containing differently expressed genes with their values of fold change were uploaded to the application.

1. Network Generation. Data sets containing gene identifiers and corresponding expression values or protein identifiers for protein interactions were uploaded respectively into in the application. Each identifier was mapped to its corresponding object in the Ingenuity® Knowledge Base. Cutoffs of differential gene expression value were not set as the identify molecules whose expression were pre-processed and the cutoffs were pre-decided before the upload. These molecules, called Network Eligible molecules, were overlaid onto a global molecular network developed from information contained in the Ingenuity Knowledge Base. Networks of Network Eligible Molecules were then algorithmically generated based on their connectivity.
2. Functional Analysis of an Entire Data Set. The Functional Analysis identified the biological functions and/or diseases that were most significant to the data set. Molecules were associated with biological functions and/or diseases in the Ingenuity Knowledge Base were considered for the analysis. Right‐tailed Fisher’s exact test was used to calculate a p‐value determining the probability that each biological function and/or disease assigned to that data set is due to chance alone.
3. Functional Analysis of a Network. The Functional Analysis of a network identified the biological functions and/or diseases that were most significant to the molecules in the network. The network molecules associated with biological functions and/or diseases in the Ingenuity Knowledge Base were considered for the analysis. Right‐tailed Fisher’s exact test was used to calculate a p‐value determining the probability that each biological function and/or disease assigned to that network is due to chance alone.
4. Canonical Pathway Analysis of Entire Data Set. Canonical pathways analysis identified the pathways from the IPA library of canonical pathways that were

most significant to the data set. Molecules from the data set were associated with a canonical pathway in the Ingenuity Knowledge Base were considered for the analysis. The significance of the association between the data set and the canonical pathway was measured in 2 ways: 1) A ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway is displayed. 2) Fisher’s exact test was used to calculate a p‑value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone.

5. Network/My Pathways/ Path Designer Graphical Representation. A network/My Pathways is a graphical representation of the molecular relationships between molecules. Molecules are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature, from a textbook, or from canonical information stored in the Ingenuity Knowledge Base. Human, mouse, and rat orthologs of a gene are stored as separate objects in the Ingenuity Knowledge Base, but are represented as a single node in the network. The intensity of the node color indicates the degree of up‑ (red) or down‑ (green) regulation. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed with various labels that describe the nature of the relationship between the nodes (e.g., P for phosphorylation, T for transcription), all of which were annotated in Appendix.

Table 2.1. bacterial strains used

Strain	Source
DH5α	Invitrogen
BL21-CodonPlus®(DE3)-RIPL cells	Stratagene
XL10-Gold	Stratagene
BL21 Gen-X Competent E.coli	Genlantis

Table 2.2. Synthetic Oligonucleotides used

Name	Reference	5'→3' Sequence
MedakaMBD4_forward	This Report	CCTGGGATCCATGAAAAATCACAAAGATGA
MedakaMBD4_reverse	This Report	CCTGCTCGAGCAGTGGGAGTGCTTCTGAGG
xlMBD4_forward	This Report	CCTGGGATCCAATGTAGATGACAAAAATTC
xlMBD4_reverse	This Report	CCTGCTCGAGGTCTCTGTCCTCATCTTCCC
XlMeCP2_forward	This Report	CCTGGGATCCAGAGACAGGGGTCCCATGTA
XlMeCP2_reverse	This Report	CCTGCTCGAGAGGCTTTTTCGGTTGCTTCT
MedakaMBD4_New_forward	This Report	CCTGGGATCCGCCTGCTTTAAAGAACTGAC
MedakaMBD4_New_reverse	This Report	CCTGCTCGAGTCTCCTCTGCTTTTGACTCA
xlMBD4_New_forward	This Report	CCTGGGATCCGAAAAGAGCCAACTACAAGA
XlMeCP2_New_forward	This Report	CCTGGGATCCGCTGTTCCTGAATCTTCTGC
mMBD4_New_forward	This Report	CCTGGGATCCGTGATAAGCAGTGAGCGCAG
mMBD4_New_reverse	This Report	CCTGCTCGAGCTGCAGGGAAGTCAGAGCTG
MedakaMBD4_stop_reverse	This Report	CCTGCTCGAGTCATCTCCTCTGCTTTTGACTCA
xlMBD4_stop_reverse	This Report	CCTGCTCGAGTCAGTCTCTGTCCTCATCTTCCC
XlMeCP2_stop_reverse	This Report	CCTGCTCGAGTCAAGGCTTTTTCGGTTGCTTCT
mMBD4_stop_reverse	This Report	CCTGCTCGAGTCACTGCAGGGAAGTCAGAGCTG
xlmbd4_full_forward	This Report	CCTGGGATCCATGGCTGCTCCTGTGCCGGA
xlmbd4_full_reverse	This Report	CCTGCTCGAGTCAGTTGAGGCCAAGCTTCTTGT
xlMBD4_27aahmecp2_forwa rd	This Report	CCTGGGATCCGCTGTGCCGGAAGCTTCTGCCTCCCCCAAACAG CGGCGCTCCATCATCCGTGACCGGGGACCCATGTATGATGACC CCACCCTCCCACATGGATGGAAAA
mMBD4_27aahmecp2_f	This Report	CCTGGGATCCGCTGTGCCGGAAGCTTCTGCCTCCCCCAAACAG CGGCGCTCCATCATCCGTGACCGGGGACCCATGTATGATGACC CCACCGTCCCGTGTGGATGGGAAA
medMBD4_27aahmecp2_f	This Report	CCTGGGATCCGCTGTGCCGGAAGCTTCTGCCTCCCCCAAACAG CGGCGCTCCATCATCCGTGACCGGGGACCCATGTATGATGACC CCACCATGCCTCCAGGCTGGATTA
xlMeCP2_27aahmecp2_f	This Report	CCTGGGATCCGCTGTGCCGGAAGCTTCTGCCTCCCCCAAACAG CGGCGCTCCATCATCCGTGACCGGGGACCCATGTATGATGACC CCACCCTTCCTGAAGGCTGGACAC
Oligos for Y2H		
mMBD4_Y2H_forward	This Report	CCTGCCATGGGTGATAAGCAGTGAGCGCAG
mMBD4_Y2H_reverse1	This Report	CCTGGGATCCCTGCAGGGAAGTCAGAGCTG
mMBD4_Y2H_reverse2	This Report	CCTGGTCGACCTGCAGGGAAGTCAGAGCTG
mMBD4_Y2H_newF	This Report	CCTGCCATGGATGGTGATAAGCAGTGAGCGCAGCTCC

mMBD4_Y2H_newR1	This Report	CCTGGGATCCTCACTGCAGGGAAGTCAGAGCTGCC
mMBD4_Y2H_newR2	This Report	CCTGGTCGACTCACTGCAGGGAAGTCAGAGCTGCC
mMBD4_Y2H_fullF1	This Report	CCTGCATATGGAGAGCCCAAACCTTGGGGACA
mMBD4_Y2H_fullR1	This Report	CCTGGTCGACTCAAGATAGACTTAATTTTTCATGA
mMBD4_Y2H_fullF2	This Report	CCTGGAATTCATGGAGAGCCCAAACCTTGGGGACA
mMBD4_Y2H_fullR2	This Report	CCTGGGATCCTCAAGATAGACTTAATTTTTCATGA
Oligos for repression test		
pcDNA31_mMBD_F	This Report	CCTGGGATCCATGGTGATAAGCAGTGAGCGCAGCTCC
pcDNA31_mMBD_R	This Report	CCTGCTCGAGTCACTGCAGGGAAGTCAGAGCTGCC
pcDNA31_mMBD4full_F1	This Report	CCTGGCTAGCATGGAGAGCCCAAACCTTGGGGAC
pcDNA31_mMBD4full_F2	This Report	CCTGGGATCCATGGAGAGCCCAAACCTTGGGGAC
pcDNA31_mMBD4full_R1	This Report	CCTGAAGCTTTCAAGATAGACTTAATTTTTCATG
pcDNA31_mMBD4full_R2	This Report	CCTGCTCGAGTCAAGATAGACTTAATTTTTCATG
pcDNA31_mMBD4full_R3	This Report	CCTGGAATTCTCAAGATAGACTTAATTTTTCATG
mMBD4_repress_fullF1	This Report	CCTGAGATCTATGGAGAGCCCAAACCTTGGGGACA
mMBD4_repress_fullR1	This Report	CCTGCTCGAGTCAAGATAGACTTAATTTTTCATGA
mMBD4_repress_fullF2	This Report	CCTGGAATTCATGGAGAGCCCAAACCTTGGGGACA
mMBD4_repress_fullR2	This Report	CCTGCTCGAGTCAAGATAGACTTAATTTTTCATGA
mMBD4_repress_mbdF1	This Report	CCTGAGATCTGTGATAAGCAGTGAGCGCAGC
mMBD4_repress_mbdR1	This Report	CCTGCTCGAGTCACTGCAGGGAAGTCAGAGCTGCC
mMBD4_repress_mbdF2	This Report	CCTGGAATTCGTGATAAGCAGTGAGCGCAGC
mMBD4_repress_mbdR2	This Report	CCTGCTCGAGTCACTGCAGGGAAGTCAGAGCTGCC
xlMBD4_repress_fullF1	This Report	CCTGAGATCTATGGCTGCTCCTGTGCCGGAGAGCCC
xlMBD4_repress_fullR1	This Report	CCTGCTCGAGTCAGTTGAGGCCAAGCTTCTTGTG
xlMBD4_repress_fullF2	This Report	CCTGGAATTCATGGCTGCTCCTGTGCCGGAGAGCCC
xlMBD4_repress_fullR2	This Report	CCTGCTCGAGTCAGTTGAGGCCAAGCTTCTTGTG
xlMBD4_repress_mbdF1	This Report	CCTGAGATCTGAAAAGAGCCAACTACAAGACAGC
xlMBD4_repress_mbdR1	This Report	CCTGCTCGAGTCAGTCTCTGTCCTCATCTTCCCCC
xlMBD4_repress_mbdF2	This Report	CCTGGAATTCGAAAAGAGCCAACTACAAGACAGC
xlMBD4_repress_mbdR2	This Report	CCTGCTCGAGTCAGTCTCTGTCCTCATCTTCCCCC
Oligos for mutagenesis		
ins_18nt_after_455	This Report	CTTTAAGAAGGAGATATACATATGCATCATCATCATCATCATC AAGTTTCAGTTGAAACCACTCAA
ins_18nt_after_455_antisense	This Report	TTGAGTGGTTTCAACTGAAACTTGATGATGATGATGATGATGC ATATGTATATCTCCTTCTTAAAG
Oligos for sequencing		
TF_forward	This Report	GGAATTCCATATGCAAGTTTCAGTTGAAAC
TF_reverse	This Report	GAGCTGATGAACCAGCAGGCGGGATCCCG
TF_right_reverse	This Report	CGGGATCCCGCCTGCTGGTTCATCAGCTC
TF_right1_175	This Report	CATAACGCTGAGCGACGATA
TF_right2_151	This Report	TCGGTACTTTGCCTTTACGG
TF_right3_69	This Report	GCTGTCAGCAGCGATTGTAA
TF_right4_391	This Report	CAACAACCGGCTTTTCAACT
mMBD4_Seq_forward	This Report	GTGATAAGCAGTGAGCGCAG

mMBD4_Seq_reverse	This Report	CTGCAGGGAAGTCAGAGCTG
mMBD4_Seq_lprimer545	This Report	CTTCAGAGTCGCCAGAAAGC
mMBD4_Seq_lprimer1056	This Report	ACAGGATGGCTCTGAAATGC
xlMBD4_116R_seq_primer	This Report	CATTGCCGTATTTGCC
xlMBD4_559R_seq_primer	This Report	GTGAAGGTATCAACGTC
xlMBD4_958R_seq_primer	This Report	CTACAAACAGTGCTTC
pcdna3.1_BGH	This Report	TAGAAGGCACAGTCGAGG
T7terminator	This Report	CTAGTTATTGCTCAGCG
T7_reversed	This Report	TAATACGACTCACTATA
Probes for EMSAs		
S1(63)	(Klose et al., 2005b)	GCAGCACGGTGGGGGGCCGGAGTTAAGGACTCGTTGTC
S2(48)	(Klose et al., 2005b)	CCAAGGTTGGTCGACCCCGGAGATAACGAAGCTCGGTG
S3(11)	(Klose et al., 2005b)	CAAGGAGGGATAGTAGCCGGACAGAAGAAATCTCTGGG
S4(60)	(Klose et al., 2005b)	GCTTGACTACCCAGACCCGGGGCGATCAATTACCGTTG
S5(83)	(Klose et al., 2005b)	CACTCTGGAAGATACGCCGGGAGTAAGCGAGCTGTAGG
U6(64B)	(Klose et al., 2005b)	GCGGTGTTAGGGTCCCCCGGTGCGCCAATGGCTGCTGA
U7(58B)	(Klose et al., 2005b)	CGCTCGGGGGGTTGGCCCGGATCCCTGGCACCTCTAGT
S1reverse	(Klose et al., 2005b)	GACAACGAGTCCTTAACTCCGGCCCCCCACCGTGCTGC
S2reverse	(Klose et al., 2005b)	CACCGAGCTTCGTTATCTCCGGGGTCGACCAACCTTGG
S3reverse	(Klose et al., 2005b)	CCCAGAGATTTCTTCTGTCCGGCTACTATCCCTCCTTG
S4reverse	(Klose et al., 2005b)	CAACGGTAATTGATCGCCCCGGGTCTGGGTAGTCAAGC
S5reverse	(Klose et al., 2005b)	CCTACAGCTCGCTTACTCCCGGCGTATCTTCCAGAGTG
U6reverse	(Klose et al., 2005b)	TCAGCAGCCATTGGCGCACCGGGGGACCCTAACACCGC
U7reverse	(Klose et al., 2005b)	ACTAGAGGTGCCAGGGATCCGGGCCAACCCCCCGAGCG
Sm probe duplex	(Prokhortchouk et al., 2001b)	CAGCAGCMGMGCCCAAMGCTGGGA
RT-PCR for checking Mbd4 and Tdg knockouts		
F_M4exon3internal	(Fiona MacDougall, 2007)	TCAAGACACGAAGCAAGTGG
R_ M4exon3internal	(Fiona MacDougall, 2007)	CCCTTTCTGTCTCCCTTCG
F_M4exon2	(Fiona MacDougall, 2007)	TACCACAGCGACAGAAGGC

R_M4exon4	(Fiona MacDougall, 2007)	CTTGTGTCCGTGGGATGC
M4JF	(Fiona MacDougall, 2007)	AAGGTGGCACCTAGAGCTGTCG
M4JR	(Fiona MacDougall, 2007)	GGATATTCGGTGCTGTCGCTCG
mTdgEx5	(Fiona MacDougall, 2007)	CGGATTCACCAACATGGTGG
mTdgEx6L	(Fiona MacDougall, 2007)	GTTTCTGCACCAGGATGCG
M4KO_F_p1	(Millar et al., 2002b)	CGGCTCAGTGGTTTGAGTTGAGTTGTGTTT
M4KO_R_p3R	(Millar et al., 2002b)	CACCTTCTGAACCACAGATGCTCTT
M4KO_p1/2	(Millar et al., 2002b)	CTTGGGATCGCTGCAAGGAAGATA
M4KO_p2	(Millar et al., 2002b)	AGTACCACAGCGACAGAAGGCCACAAG

For Mass Spec

hMBD4mbd_mass_F	This Report	CCTGGAATTCATGAAAAGAAGCAGTGAATGTAA
hMBD4mbd_mass_R	This Report	CCTGGCGGCCGCTAGATGGGATGTCAGG
hMBD4full_mass_F	This Report	CCTGGAATTCATGGGCACGACTGGGCTGGAGAG
hMBD4full_mass_R	This Report	CCTGGCGGCCGCAGATAGACTTAATTTTTCATG
mMBD4mbd_mass_F	This Report	CCTGGCTAGCGTGATAAGCAGTGAGCGCAG
mMBD4full_mass_F	This Report	CCTGGCTAGCGAGAGCCCAAACCTTGGGGA
mMBD4full_mass_R	This Report	CCTGCTCGAGTCATCAAGATAGACTTAATTTTTC

For MBD4 interaction domain

hMBD4_F_1_156	This Report	CCTGGGATCCggcacgactgggctggagag
hMBD4_R_1_156	This Report	CCTGCTCGAGtttatatcttgacttgatac
hMBD4_R_1_408	This Report	CCTGCTCGAGctgtgttcgtgggatggtat
hMBD4_F_408_580	This Report	CCTGGGATCCcagatagaaagaaggaaaac
hMBD4_R_408_580	This Report	CCTGCTCGAGagatagacttaatttttcat
hMBD4_F_156_455	This Report	CCTGGGATCCaaagactgcagcatggcagc
hMBD4_R_156_455	This Report	CCTGCTCGAGatcatgaaaaagtgtttctt

Primers used for expression analysis from cDNA (all are intergenic)

RT_Igfbp4_F	This Report	GAGAAGCCCCTGCGTACAT
RT_Igfbp4_R	This Report	GTTGGGATGTTCGCTCTCAT
RT_Gfpt1_F	This Report	GAGCTCACCCTCGCTTCTG
RT_Gfpt1_R	This Report	CCAAGATTTCTCGTCTTGTTCG
RT_ Cryl1_F	This Report	GGGGTGGTGATCGTAGGC
RT_ Cryl1_R	This Report	TTTCCAGGGCGTCTGTTATC
RT_ Dio3_F	This Report	AGGTTGGTGGTCGGAGAAG
RT_ Dio3_R	This Report	AACAGCACGAGGCACGAG
RT_ Odz4_F	This Report	GTACTGGAGCCTGGGAACC
RT_ Odz4_R	This Report	TCTCTCAGCCTCAGACTGTCC
RT_ Sdpr_F	This Report	TGTCCAGGTGAAGAGACTGG
RT_ Sdpr_R	This Report	TGGCTCCTTCACAAACACAC
RT_ Wnt4_F	This Report	CCTGCGACTCCTCGTCTTC
RT_ Wnt4_R	This Report	CTCTGGATCAGGCCTTTGAG
RT_ C1qtnf3_F	This Report	TTGCTTTTCCTCCCATTTTG
RT_ C1qtnf3_R	This Report	CCTTGGTAACCACGAAATCC
RT_ Anxa4_F	This Report	AGACTGGGCGGGTCCTATC
RT_ Anxa4_R	This Report	CATGCTGTAGCTCACGTTCAG

RT_ Cp_F	This Report	AGCCTTGGCAAGAGATAAGC
RT_ Cp_R	This Report	CGATCTGGACCATTTTGAAG
RT_ Emb_F	This Report	TACCCGCTCCTCCTTCTATG
RT_ Emb_R	This Report	AAGGTTTGAGTATTTCGCCATC
RT_ Gjb2_F	This Report	TTCAGACCTGCTCCTTACCG
RT_ Gjb2_R	This Report	GAGGATGCTCTGGAGTGTGC
RT_ Ifitm1_F	This Report	GGATCCTAGGCCTTTCCATC
RT_ Ifitm1_R	This Report	TCGCAGAAGTTGAGATGTGG
RT_ Igfbp5_F	This Report	TTTGCCTCAACGAAAAGAGC
RT_ Igfbp5_R	This Report	GAAGACCTTGGGGGAGTAGG
RT_ Il17d(1-2)_F	This Report	CAAGTCTGGAAAGCATCACG
RT_ Il17d(1-2)_R	This Report	GAGCTGCAGCGTGTGGTG
RT_ Mtap2_F	This Report	TTCTTTCCCTCGTTTCTTCG
RT_ Mtap2_R	This Report	GAAACTTGGAGCACACAGCA
RT_ Pcsk6_F	This Report	CGCACGGCTACCTCAACT
RT_ Pcsk6_R	This Report	GGGTCCATTCTGAGGAAGGT
RT_ Tox_F	This Report	GCTTGATGTGAGAGTGAAATGG
RT_ Tox_R	This Report	CATATACATGTTCTCCCCGTCA

PCR primers

Custom oligonucleotides for PCR or DNA sequencing were purchased from Sigma- Aldrich. An oligonucleotide stock solution was prepared by the resuspension of lyophilised oligonucleotides in dH2O to 100pmol/µl. This stock solution was diluted in dH2O to give a 25pmol/µl PCR primer solution or a 1pmol/µl DNA sequencing primer solution. Resuspended oligonucleotides were stored at −20°C.

Oligonucleotides used as EMSA assay substrates

Custom oligonucleotides for use as EMSA assay substrates were purchased from Sigma- Aldrich. An oligonucleotide stock solution was prepared the same as PCR primers above.

Table 2.3. Plasmids used

Plasmid	Selectable markers	Reference/Source
EMSA		
pET21b(-)-TF-UB-single6histag (at downstream)	Amp	Thapa, et al. 2008
pET21b(-)-6histag-TF-UB_6histag (atupstream and downstream)	Amp	This Report
pHUE	Amp	Baker, et al. 2005

pHUsp2-cc	Amp	Baker, et al. 2005
pHUE-H6-Ub-M-GSTP1	Amp	Baker, et al. 2005
For Y2H		
pGBKT7	Kan	Clontech
MBD Domain Cloning		
medaka	Amp	Ruzov, et al. 2009
xlMBD4	Amp	Ruzov, et al. 2009
xlMeCP2	Amp	Ruzov, et al. 2009
mMBD4	Amp	Ruzov, et al. 2009
Repression assay		
PCS2p+eGFPbgl2seq	Amp	This Report
PCS2p+nls_eGFPbgl2seq	Amp	This Report
pGL3-b-Tex19(+224,-56)	Amp	This Report
TK-Renilla	Amp	Promega
pCDNA3.1(+)	Amp	Molecularlab
IOM20218-pDEST26	Amp	imaGenes
IOM20218-pDEST27	Amp	imaGenes
IOM20218-pdEYFP-C1amp	Amp	imaGenes
IOM20218-pT-Rex-DEST31	Amp	imaGenes
pGemT-Easy	Amp	Promega

Table 2.4. Cell culture conditions and genotypes.

Cell type	Genotype	Culture Media	Passaged
MEF	$p53^{-/-}$	DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
pMEF	WT	DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 6
293T	WT	DMEM; 10% FCS; 1000U/ml Pen; 650µg/ml Strep	1 in 10
CMT93	WT	DMEM; 10% FCS; 1000U/ml Pen; 650µg/ml Strep	1 in 10
MEF	$Mbd4^{-/-}$	DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 10 (virus transformed); 1 in 4 (non-transformed)
ES	WT, E14	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5 /ml gelatin

Chapter 3: Identification of candidate protein partners of MBD4 by high throughput yeast two-hybrid screens and affinity purification protein identification

3.1. Introduction

MBD4 has been linked to the mismatch repair pathway as part of a demethylation process (It has been implicated as a glycosylase to excise the thymine mismatches generated by deaminases); apoptosis signalling and transcriptional repression (Meng et al., 2011). Most of these findings started from insights of MBD4 interacting proteins and the relevant potential functions of MBD4 were then confirmed and further investigated. For example, a yeast two-hybrid screen (Y2H) using MLH1 as bait identified MBD4 as candidate interacting protein and the subsequent validations and functional investigations confirmed that MBD4 is involved in the mismatch repair pathway (MMR) (Bellacosa et al., 1999). Another example comes from co-immunoprecipitation of MBD4, where several deaminases such as AID and APOBECs were suggested to co-immunoprecipitate with MBD4 in zebrafish tissues, and the further investigations implicated MBD4 as a possible component of the demethylation process in zebrafish (Rai et al., 2008). MBD4 has a fusion-like domain structure, with an MBD domain and a glycosylase domain separated by a 'desert' region that exhibits poor conservation of amino acid sequence (Figure. 3.1). The fusion-like domain property of MBD4 implies that it may perform functions in addition to its activity as a glycosylase. It is noteworthy that there were published Y2H studies using MBD4 as bait, however, only a few candidates were identified (Millar, 2002), most of which were related to the glycosylase/repair function of MBD4. This may be due to a limitation of the cDNA library and low-sensitivity Y2H system that was used. For example, Millar chose a mouse brain cDNA library as the pool for prey (Millar, 2002), but the expression of MBD4 in brain is very low (based on the BioGPS, www.biogps.org). This suggests that the library is not very suitable and that MBD4 interacting proteins that are not expressed in brain could be missed. Most of the known MBD4 interacting proteins bind at the C-terminal domain of MBD4 (Figure. 3.1), with little known about the N-terminal regulatory part of MBD4. Our lab has previously identified DNMT1 as a key partner of MBD4, and showed that the N-terminal portion containing the MBD domain of MBD4 can account for the association (Ruzov et al., 2009). In addition, several other MBD4 partners that interact with MBD4 at its N-terminal domain have been identified by affinity purification of MBD4 complex, including the afore mentioned deaminases and several transcriptional repressors including SIN3a and HDAC1 (Fukushige et al., 2006, Kondo et al., 2005) (Figure. 3.1). To date, a systematic screen for MBD4 interacting protein partners is lacking, which could lead to a better understanding of the genetic and epigenetic functions of the MBD4 protein and its related machinery.

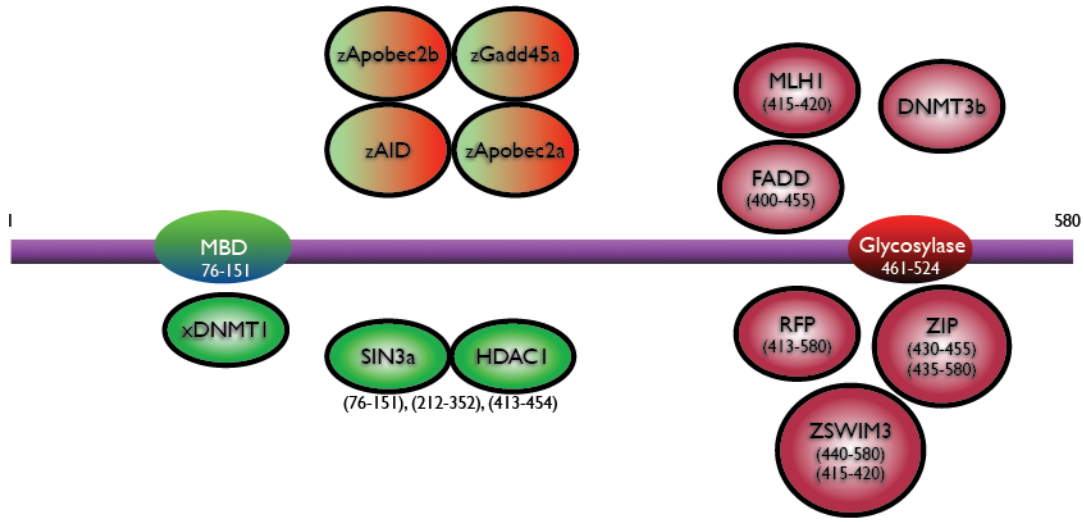


Figure. 3.1. MBD4 interacting proteins from literature review. If known, the minimum region of MBD4 involved in the interaction is shown below each protein name. The proteins in green interact with MBD4 at its N-terminal MBD domain or its intervening region. The proteins in red associate with MBD4 at or near its C-terminal glycosylase domain.

High-throughput methods such as Y2H screening and affinity purification of complexes followed by mass spectrometry analysis (AP/MS) are the two most frequently used methods for interactomics research (Brückner et al., 2009). Y2H is a well-established genetic in vivo approach that allows detection of interacting proteins in living yeast cells (Fields and Song, 1989). In Y2H system, interaction of bait protein and prey protein activates delicately designed reporter genes, which allow mated yeasts to be screened by either chemical or colour selection. Y2H can be performed with high-throughput on a genome-wide scale, and the expense is relatively low compared with an equivalent AP/MS based method (Brückner et al., 2009). More recently AP/MS has become a powerful tool for high-throughput screening of protein interactions (Gloeckner et al., 2007). MS coupled to single- or double affinity purification is now routinely applied to identify proteolytic fragments of proteins or protein complexes (Fenn et al., 1989). Single step purification such as Flag- and Strep- tag can be used for affinity purifications. However, two-step purification using the TAP (tandem affinity purification) approach was shown to be more efficient (Gloeckner et al., 2007). The idea of TAP is that it contains a fusion of two affinity tags partitioned by a protease cleavage site allowing each tag to be used sequentially to purify the protein of interest and interacting proteins (Gloeckner et al., 2007). Followed by computational analysis of MS data, the background/noise can be dramatically reduced compared to single tag purification (Brückner et al., 2009).

Y2H and AP/MS approaches detect complementary kinds of protein interactors (Brückner et al., 2009). Both approaches may detect binary direct interactions. However, Y2H may detect transient and very weak interaction that may be missed in AP/MS, while AP/MS may determine all the components of a large complex, which may not be direct interactions, and thus could not be identified by Y2H. Both screens may have their own limitations and advantages, in terms of their coverage, false-positives, and sensitivities. It is noteworthy that a combination of both approaches followed by well-designed bioinformatics analysis may lead to a fairly complete interactome data and may therefore facilitate a better biological understanding of the targeted proteins.

To explore the functional repertoire of MBD4 I decided to undertake multiple protein interaction studies to determine its potential partner proteins. I performed Y2H with an 11.5-day mouse embryonic cDNA library and multiple AP/MS experiments using tagged versions of MBD4 that were over-expressed in human cell lines. I detected ~ 380 potential interacting candidates with these assays. A significant number of candidates were detected in both assay systems. In general, further validations and functional studies of these candidates demonstrated the robustness of my initial screens, and suggest that MBD4 may impact several transcriptional and epigenetic networks along with a number of nuclear pathways that include DNA repair, transcriptional repression through histone and chromatin remodellers, and possibly RNA processing.

3.2. Results

3.2.1. Yeast two-hybrid screens with full length and MBD domain of MBD4

Most of the previously suggested MBD4 interacting proteins interact with MBD4 at or near its C-terminal glycosylase domain. Specifically, there are ~30-50 amino acids immediately upstream of its C-terminal glycosylase domain identified as the minimum interacting region for most of the previously suggested MBD4 interactions. It has been previously shown that MBD4 interacts directly with both DNMT1 and MLH1, leading to recruitment of all three proteins at DNA damage sites (Ruzov et al., 2009). The region that is thought to be responsible for the interaction of MBD4 and DNMT1 within MBD4 is its N-terminus MBD domain (Ruzov et al., 2009), which suggests that the MBD domain of MBD4 may act as an interaction domain in addition to its methyl-CpG binding capability. Thus, initially I hypothesised that MBD4 may possess additional functions that are independent of its role as

a glycosylase DNA repair protein. The MBD domain of MBD4 may be a region that recruit other partners. Herein I started to perform yeast two-hybrid screens using the MBD domain and full length Mbd4 protein as baits.

I chose to work with Clontech Matchmaker™ library construction & screening system with a mouse 11.5 days embryo cDNA library. The reason why I used this library is that the 11.5 days mouse embryo has passed the early developmental stage which MBD4 may have less significance as the mice lacking Mbd4 are viable, while the library has enriched cDNAs representing the ongoing development of the circulatory system, intestinal tract, urogenital system, central nervous system and external form (The Edinburgh Mouse Atlas, http://www.emouseatlas.org/). MBD4 is expressed ubiquitously during the mouse development and different tissues and organs, suggesting this 11.5 days mouse embryo cDNA of late developmental stage might give me potential hits in addition to what has been done previously with brain library. I cloned full-length mouse *Mbd4* and the MBD domain of mouse *Mbd4* into the bait vector, pGBKT7. The fusion bait vectors were then transformed into yeast-109 strain. The pre-transformed 11.5 days mouse embryo cDNA library for the screen was pre-transformed into yeast-187 stain and was commercially available (Clontech). The bait protein was expressed as a fusion to the Gal4 DNA-binding domain (DNA-BD), while prey proteins in the pre-transformed library are expressed as fusions to the Gal4 activation domain (AD). When bait and prey proteins interact, the transcription of four reporter genes (*ADE2*, *HIS3*, *MEL1*, and *LacZ*) can be activated, as the Gal4 DNA-BD and AD are brought into close proximity. In addition, the bait vector pGBKT7 has *TRP1* sequence, whereas the library vector has *LEU2* sequence. This means in theory only diploid cells containing bait and prey of positive interacting candidate could grow on the selection plates lacking histidine, adenine, leucine and tryptophan. They also express a-galactosidase (*MEL1*), turning blue in the presence of the chromagenic substrate X-α-Gal.

I performed control experiments first, and confirmed there were no autoactivation and toxicity in my baits. I chose quadruple-dropout (SD/-Ade/-His/-Leu/-Trp/X-α-Gal), the most stringent selection, for my initial screen, and followed a modified mating protocol to increase the mating efficiency (Figure. 3.2). Over 2000 positive colonies of MBD4 full length screen grew and turned blue on selection plates, and ~150 positive colonies from

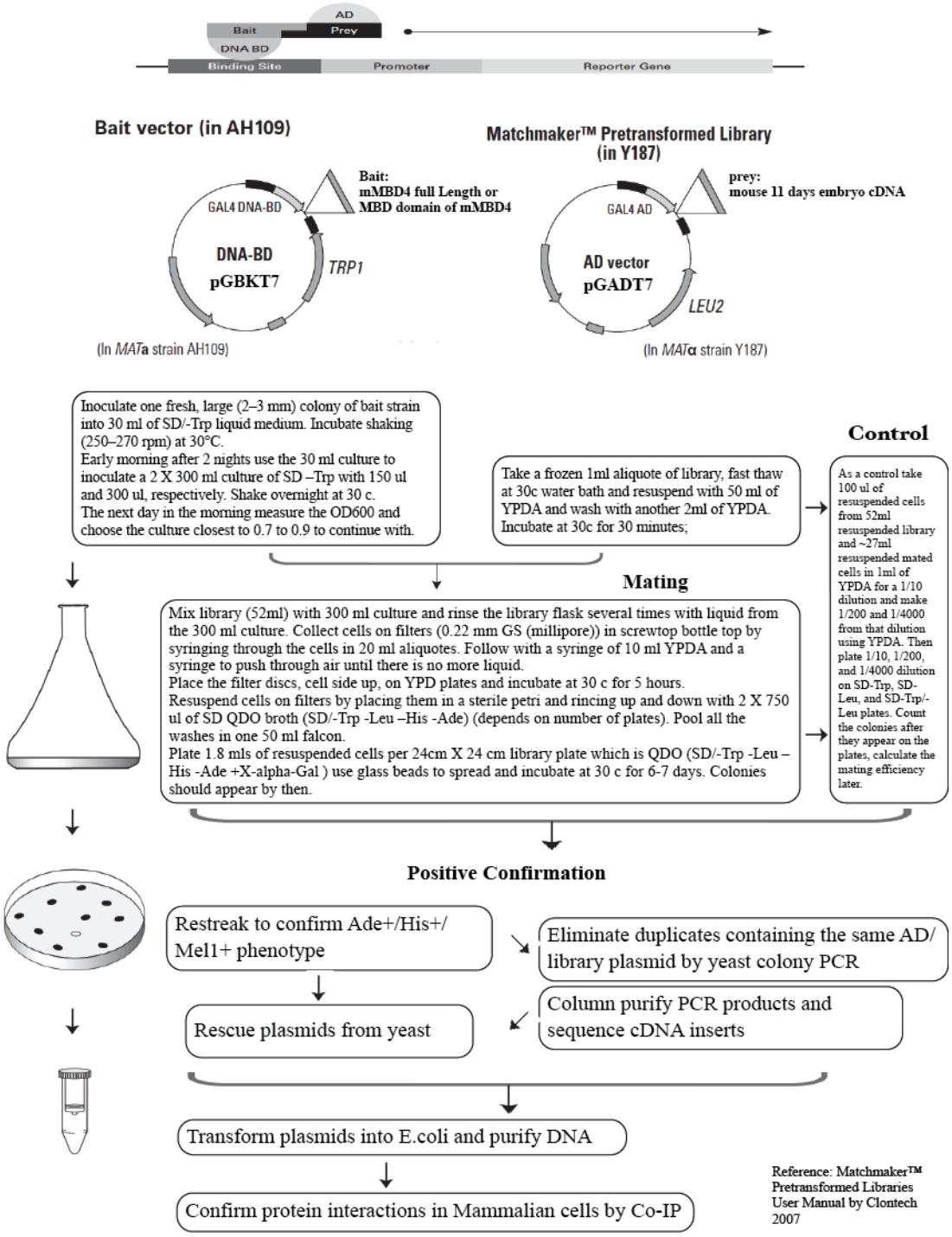


Figure. 3.2. Yeast two-hybrid flow chart for detecting MBD4's potential interacting partners. Picture parts are adapted from the reference indicated.

MBD domain screen. I next performed quality control and identification of cDNA hits. The positive colonies showing deepest blue (1440 from MBD4 full length screen, 144 from MBD domain screen) were picked out, and were restreaked onto fresh quadruple-dropout selection plates for further confirmation of their phenotypes (Figure. 3.3.1). The restreaking can

eliminate the false positive colonies on the initial selection plates to increase the chance of rescuing the positive prey plasmid, as there may be a mixture of blue and white colonies indicating segregation of positive interactors (blue) from non-interactors (white). Colonies that have been confirmed by restreaking were picked to perform yeast colony PCRs. I decided to directly purify the PCR products and perform sequencing to identify the inserts. The majority of interaction candidates were identified by this colony PCR-purification-sequencing approach. Plasmid-rescue was performed for the positive colonies that failed to be sequenced (n=257 (18%) in full length MBD4 screen; n=17 (11%) in MBD domain screen) (Figure. 3.3.2). These rescued plasmids were then sequenced. Using a combination of these methods, prey cDNAs were identified from all positive colonies (Figure. 3.3.2). The cDNA inserts were identified using the sequence search module of Geneious Pro software. In total, 230 individual candidates were identified using full-length mouse MBD4 as bait (Figure. 3.4) and 29 candidates were identified using the MBD domain as bait (Figure. 3.5). Of the 29 candidates found using the MBD domain, 20 candidates (69%) overlap with the full-length screen (Suppl Table. 3.1.1 & 3.1.2). This high percentage of overlapping between the two Y2H screens demonstrates the technical robustness of the assay. Moreover, Many of the candidates received multiple hits, which implies robustness of the screens. For example, the top candidate, SRSF2 (also known as SC35) has 116 hits in the MBD4 full length screen, as well as 41 hits in the MBD domain screen, suggesting it is a suitable candidate for the further validation (Figure. 3.4 & Figure. 3.5). Lastly, I performed a functional annotation of these candidates by comparison of both full length and MBD domain screens, and annotated them into functional groups. Instead of annotating these candidates using GO terms tool for enrichments, I manually classified them into functional groups based on their molecular function and biological process in the latest curated Panther classification system. This avoids the well-known lack of non-specific annotations in the GO term annotations, as the proteins in the Panther system are classified by expert biologists by functions, enabling me to look in-depth at the candidates. Interestingly, the proportions of functional groups of MBD domain and full-length screens are very similar (Figure. 3.6), and the top biological functions include signalling and transcriptional regulation, suggesting that the MBD domain of MBD4 might have a role that is independent from its glycosylase domain through its association proteins. In total over 3000 colony PCRs followed by sequencing, and over 500 plasmid rescues were performed and 239 candidates were identified.

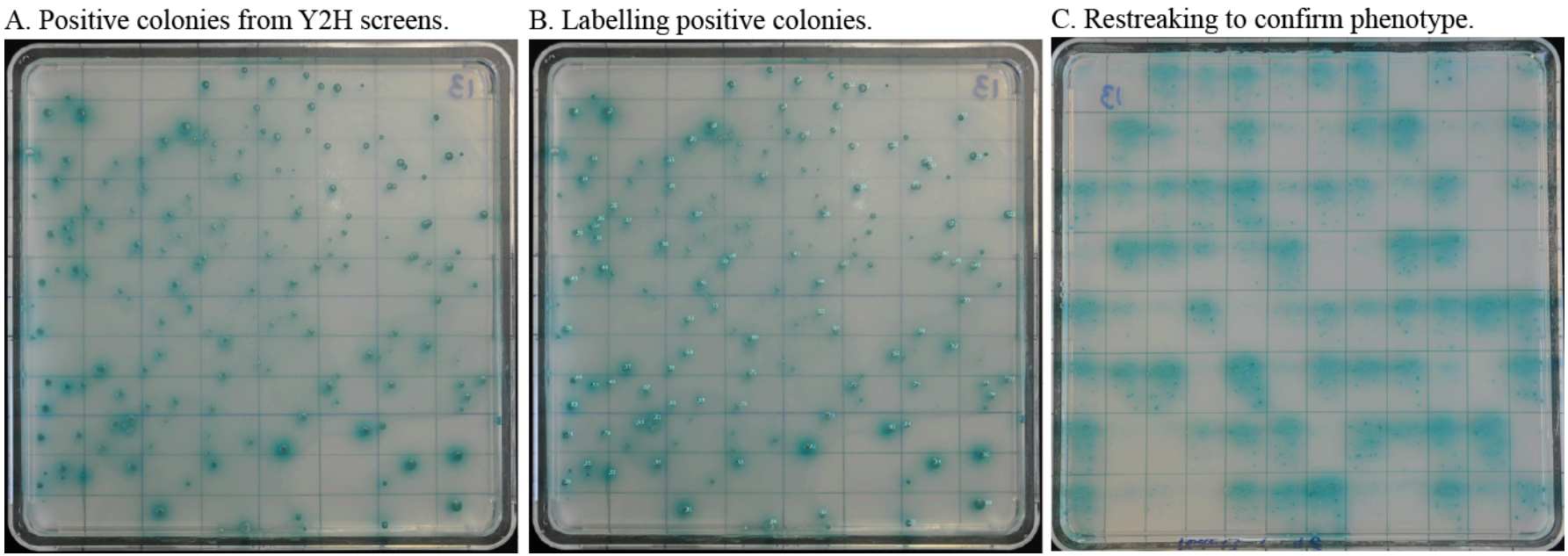


Figure. 3.3.1. The phenotype of positive colonies from Y2H screens. The colonies showing deep blue are the putative positives containing MBD4 bait and the novel interacting preys. One representative plate is shown as an example. (15 plates for each screen, 1584 colonies on 30 plates were restreaked in total)

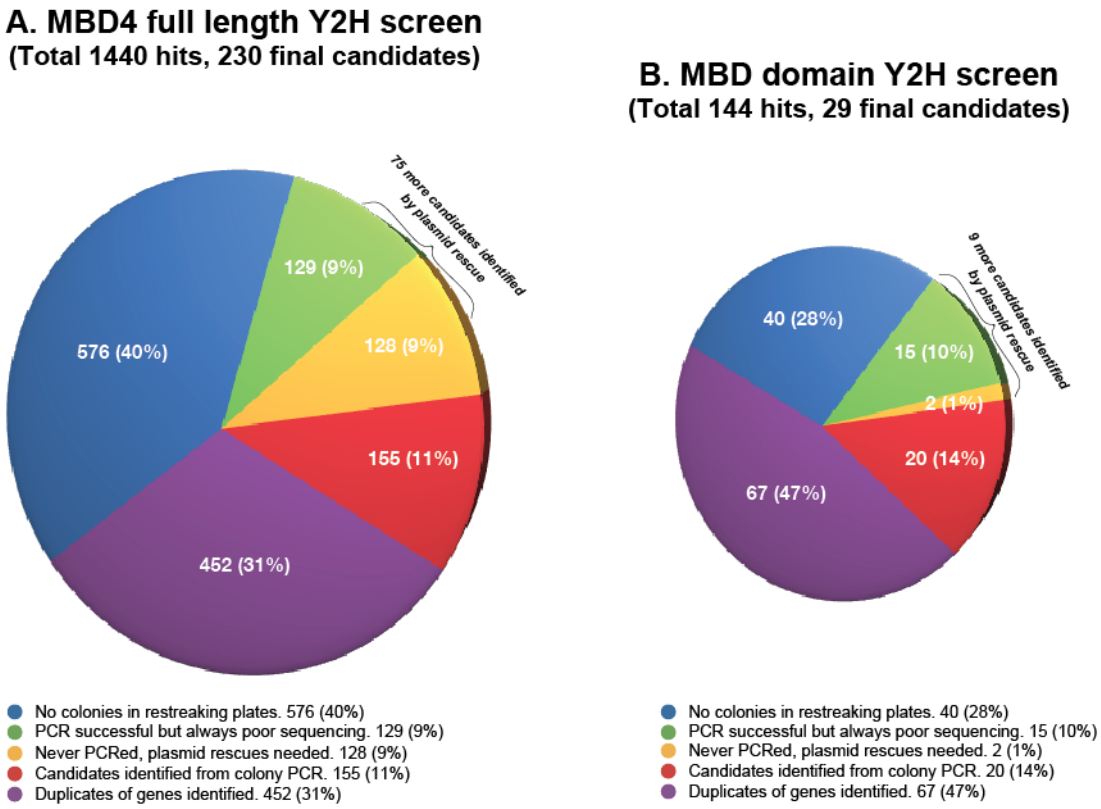
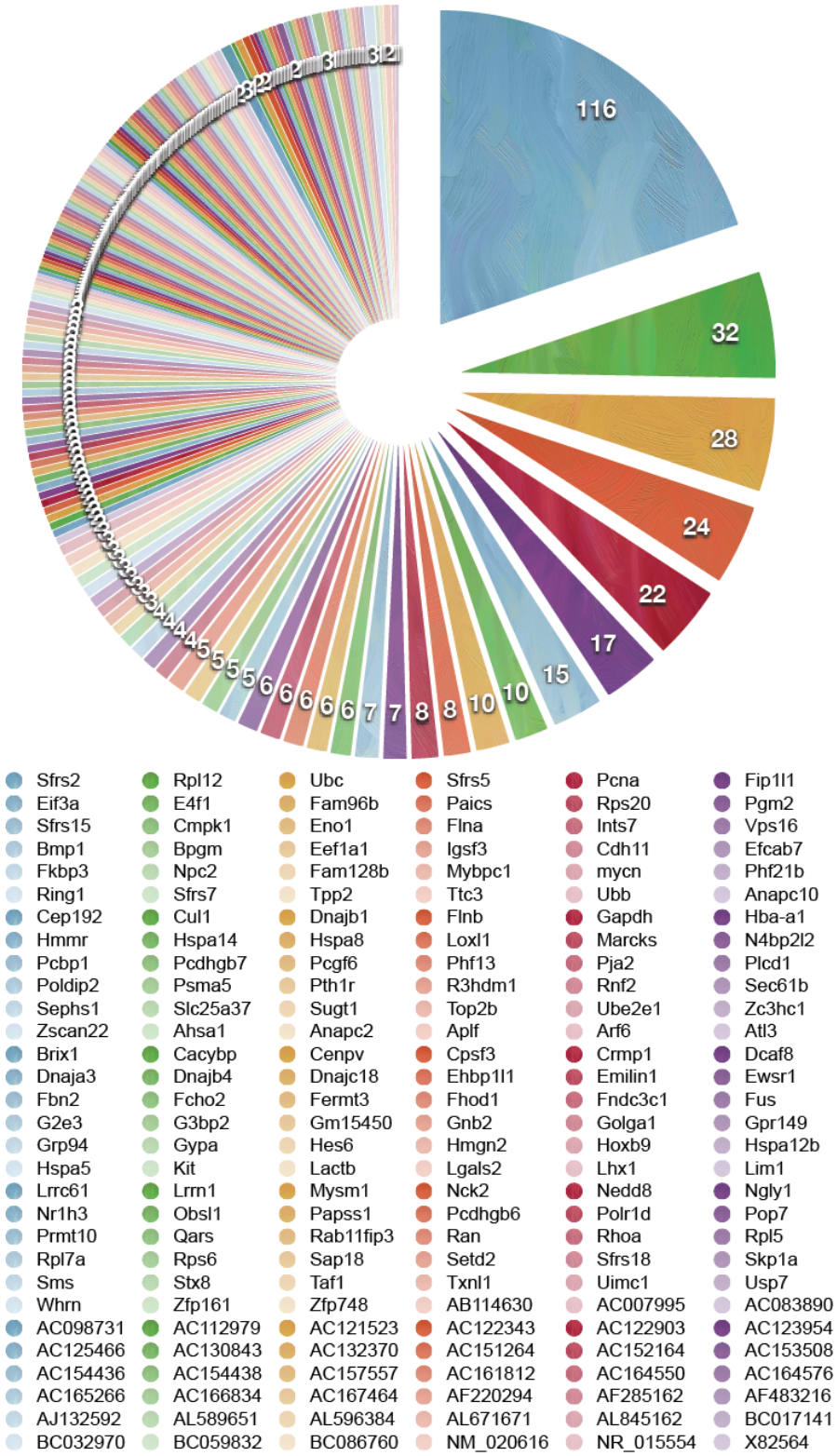


Figure. 3.3.2. Quality control and identification of Y2H hits. There are in total 1440 positive colonies selected from the MBD4 full-length screen, and 144 from the MBD domain only screen, all of which showed blue phenotype on the initial selection plates. Many colonies did not grow when re-streaked on selective plates (blue segments) and therefore were excluded. The remaining colonies were processed by colony PCRs followed by sequencing. Most of the candidates were successfully identified using this method (red and purple segments). The colonies that could not identified using this method were subjected to plasmid rescue (green and yellow segments). In total, 230 individual candidates were identified from the MBD4 full-length screen, and 29 from the MBD domain screen.


116
32
28
24
22
17
15
10
10
8
8
7
7
6
6
6
6
6
5
5
5
4
4
4
3
3
3
3
3
2
2
2

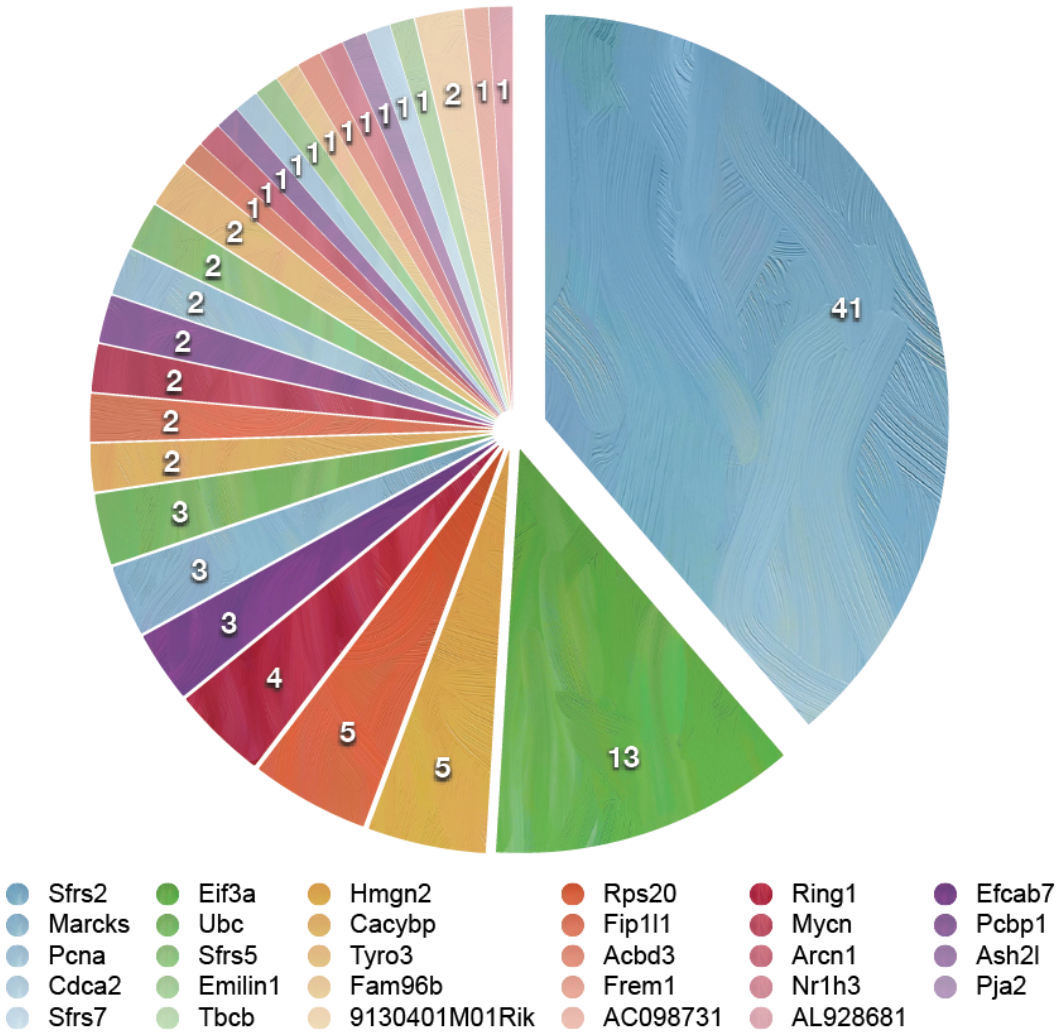


Figure. 3.5. Interacting candidates (gene names) from MBD domain Y2H screen. The numbers on the proportional pie chart indicate the number of hits received for each prey protein. The genes indicated only by GenBank accession ID are the ones that have no curated gene names.

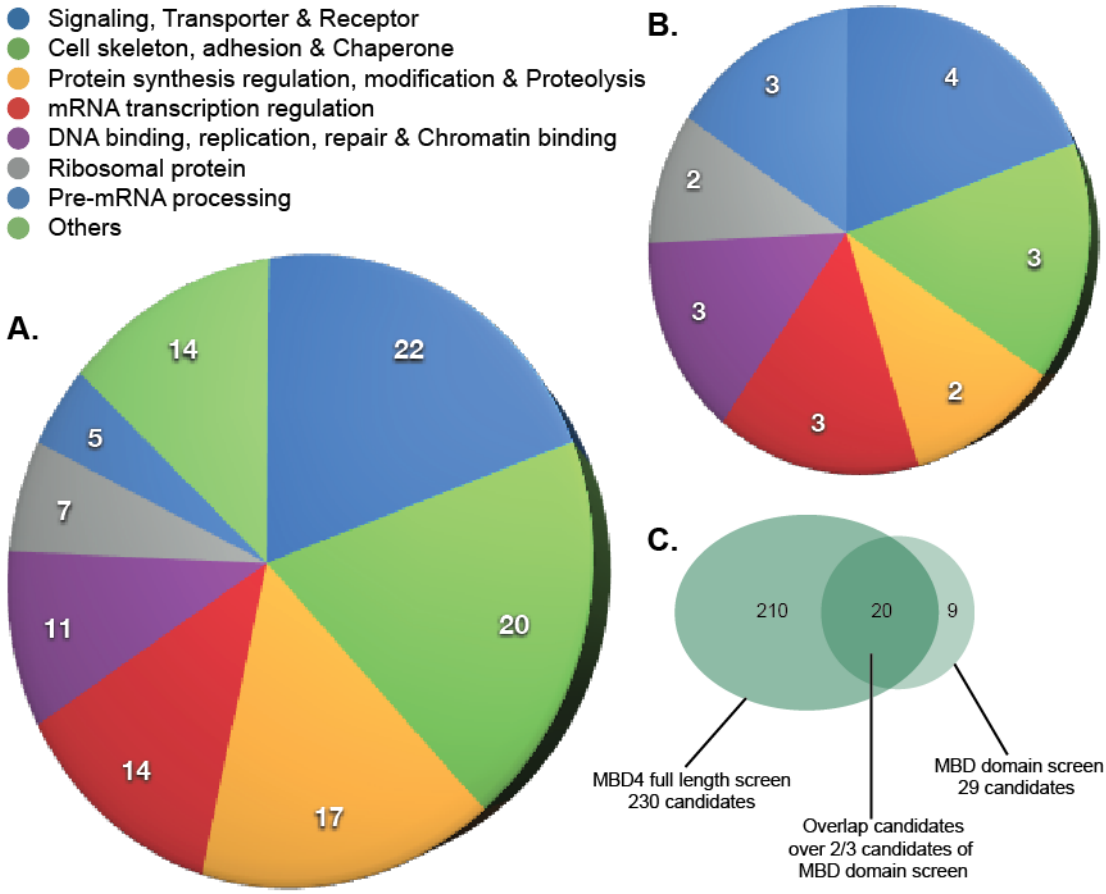


Figure. 3.6. Functional annotations of MBD4 interacting candidates from Y2H. The candidates from both Mbd4 full length Screen (A) and MBD domain screen (B) were manually curated into functional classes based on the Panther classification system. The proportions of each functional group in both screens are similar, and MBD4 seems to impact a number of pathways. (C)A venn diagram showing that most of the candidates from the MBD domain screen were overlapping with the ones from MBD4 full-length screen.

3.2.2. Identification of MBD4-associated proteins by co-immunoprecipitation followed by MS analysis

The large number of potential interacting candidates of MBD4 from my Y2H screens representing a diverse of pathways was unexpected, as MBD4 was thought mainly as a repair glycosylase protein. The large number of candidates with the possible functional diversity indicate that MBD4 might involve in a number of nuclear pathways. An immediate question was which candidates I should choose for further validation. Affinity purification followed by mass spectrometry (AP/MS) is a quick and efficient way to identify associated proteins

(Brückner et al., 2009). Moreover, AP/MS is a complementary approach to Y2H for identifying the protein complexes that MBD4 associates with (Brückner et al., 2009). Additionally the complete datasets from both high-throughput screens enabled accurate functional annotation and network analysis, as the protein-protein databases include curated protein associations from both assays of direct interactions and protein complexes.

I chose to undertake multiple co-immunoprecipitations (co-IPs) followed by MS analyses using HEK293T cells as they show high transfection efficiency using standard DNA delivery techniques (Abramson et al., 2010). As discussed in section 3.1, the development of the tandem affinity purification (TAP) technique increases the efficiency and accuracy of large-scale purification of native protein complexes. I decided to use a TAP vector called SF-TAP, which combines a tandem Strep-tag II with a FLAG-tag (Gloeckner et al., 2007). The advantage of this system is that the size of the TAP (SF-TAP) tag is reduced to 4.6 kDa, which is very efficient for cleavage, in comparison with the original 21-kDa TAP tag (Gloeckner et al., 2007). I cloned human full length *MBD4* into SF-TAP, and overexpressed this fusion protein in HEK293T cells. Initially, I followed a two-step protocol; the SF-tagged MBD4 and SF-TAP tag protein alone as control were eluted under native conditions using desthiobiotin in the first step and the FLAG octapeptide in the second step. Bound proteins were then loaded on SDS-PAGE gels and stained with coomassie blue. In TAP-tagged MBD4-expressing lanes, more bands were visible than the corresponding control lanes (Figure 3.7 B, C and E). The strongest band was observed in the TAP-tagged MBD4 expressing lane at expected size of 80–85 kDa (arrows in Figure 3.7 B, C and E) and absent from the control lane. This suggests that TAP-tagged MBD4 was successfully immunoprecipitated together with a number of partner proteins. Gel slices containing individual bands showing a clear difference in the presence or absence of MBD4 were excised from both MBD4 expressing lanes and the corresponding control lane (Figure. 3.7 B C E brackets and numbers indicated), and were sent to St. Andrews Mass Spec Services for MS analysis. Each gel slice identified several protein candidates. Importantly, MBD4 was the most abundant protein from the 80–85 kDa band, an indication that the approach was successful. The two-step TAP approach resulted in 4 robust hits including PRMT5 (Figure. 3.7 E), however, it may also remove many real hits due to the stringency of two IPs. As I have got a cohort of candidates from Y2H screens, I decided to perform single Flag- and single Strep- IP for additional screens, to inclease the IP screen coverage for overlapping candidates in both Y2H and AP/MS screens.

The combined raw data sets of the three single- and double- tagged experiments yielded a large number of putative MBD4-associated protein candidates (192 in total). To

focus on the most robust candidates and to eliminate false positives, I removed the proteins that were also detected in the relevant control lane as well as the proteins that were corresponding to known contaminants/false positives (Abramson et al., 2010). Removing these proteins yielded ~150 candidates of MBD4 interacting partners from the AP/MS screens (Suppl Table 3.2. ~150 candidates). Ten (6.7%) of these candidates overlap with Y2H screen. In the MS analysis, the MudPIT (Multidimensional Protein Identification Technology) scoring is the standard protein scoring method which calculates the sum of the ions scores but exclude the scores for duplicate matches, while the matched query number indicates how many non-redundant peptides identified from MS match in the peptide databases (Yates et al., 2009). Among the MBD4 associated candidates from the AP/MS, I identified 50 protein candidates that stood out with both high-confident MudPIT scores and the significant number of queries matched (Figure. 3.8). These 50 significant candidates did not only have significant MudPIT scores (Figure. 3.8 all 150 candidates, upper gray bar chart), but also were detected by more matched queries than their similar MudPIT neighbours (Figure. 3.8 lower blue bar chart). They are indicated in their names and light gray bars in the middle of Figure. 3.8.

The large number of candidates (93.3%) did not overlap with the Y2H screen. This is not surprising as Y2H and AP/MS screens detect complementary kind of protein interactors as discussed in the introduction section, as well as the system diversity that would be expected between the mouse embryo cDNA library and IP host HEK293T kidney derived cells. The robustness of these putative associations was then verified by validations of selected candidates by reciprocal co-IPs and cellular co-localisation experiments that will be discussed in Chapter 4. An immediate question was whether this set of proteins could imply significant functional pathways that MBD4 may impact upon, which would be testable direction(s) for further investigation?

3.2.3. Functional clustering of putative MBD4-interacting proteins by systemic network analysis

By the complementary high-throughput screens of Y2H and AP/MS, I have identified a large set of potential Mbd4 interacting candidates (n=380), including its possible direct partners and protein complex components. The high-percentage of positives in my further validation in Chapter 4 and the hits of previously suggested Mbd4 interacting proteins in my screen such as DNMT1, PCNA and MSH2 suggest that my screens are capable of detecting real

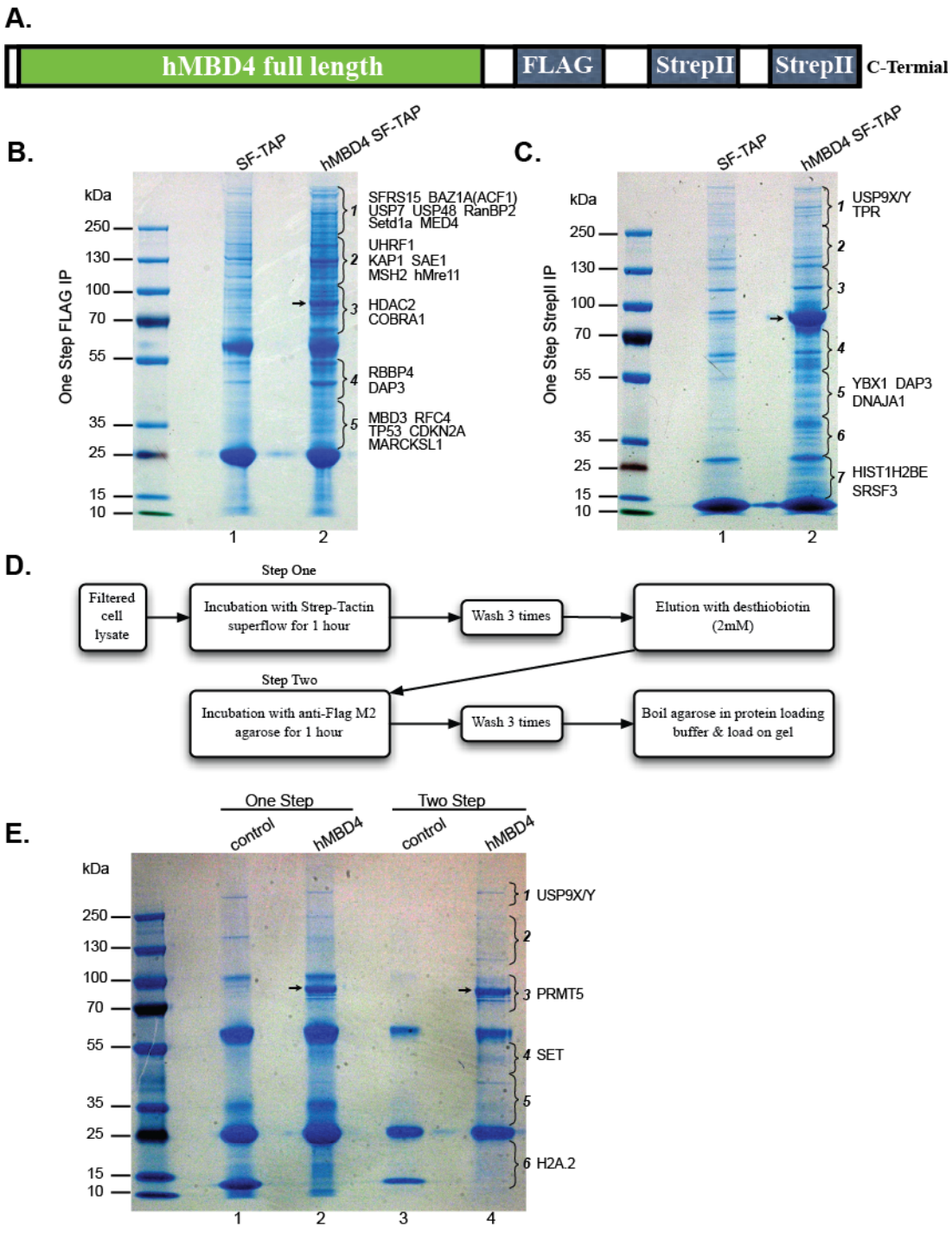


Figure. 3.7. Identification of Mbd4 -associated proteins by IP/MS screens. (A) The schematic representation of the human MBD4 full length protein fused with SF-TAP at its *C-terminal*. (B) One step purification using FLAG-tag, or (C) Strep-tag II. (D) Flow chart of the SF-TAP procedure. (E) Two step purification. Arrows in (B) (C) (E) indicate the MBD4 bands. The representative protein candidates identified in each fraction by mass spectrometry are indicated.

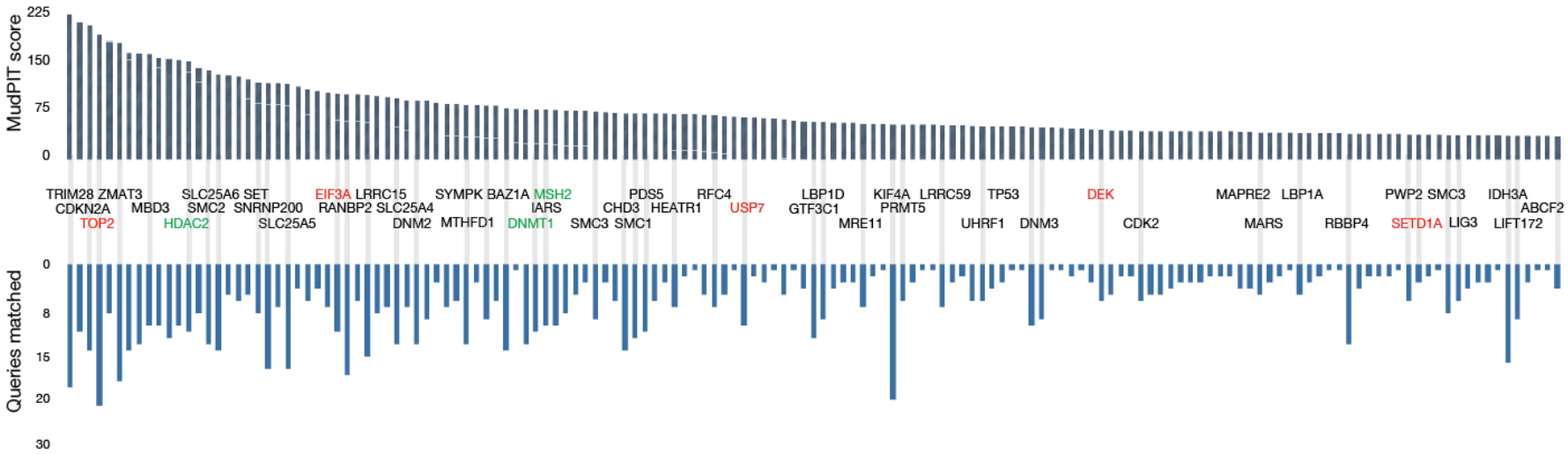


Figure. 3.8. Identification of candidate proteins from IP/MS screen. The upper bar chart represents the MudPIT scores, and all of these selected candidate genes are with a confident MudPIT score from ~40-~200. (In the MS results, the MudPIT (Multidimensional Protein Identification Technology)(Yates et al., 2009) scoring is the standard protein scoring which calculate the sum of the ions scores but exclude the scores for duplicate matches, while the matched query number indicates how many non-redundant peptides identified from MS match in the peptide databases). The lower bar chart shows how many fragment queries matched in the databases. Strongly significant candidates from MS have a significant score as well as multiple times of queries matched. Proteins labeled color red represents the ones that overlap with Y2H, green for known MBD4 interacting proteins.

MBD4 associations. Many of the MBD4-associated candidates were indeed previously reported to interact with each other. For example, it has been suggested that USP7 and UHRF1 interact with and control ubiquitination and stability of the maintenance DNA methyltransferase DNMT1 (Felle et al., 2011, Khoronenkova et al., 2011, Du et al., 2010). Interestingly, the former two proteins USP7 and UHRF1 were also identified from my screens and we have previously shown DNMT1 interacts with MBD4 directly (Ruzov et al., 2009). I mapped all the protein candidates into existing protein-protein interaction databases, prioritised the candidates based on proximity and identified the most valuable ones and/or the central hub protein candidates for the further investigations. However, my candidate list will inevitably contain a large number of false positives. The Ingenuity knowledge base is a repository of biological interactions and functional annotations created from manually curated and individually modelled relationships between proteins, genes, complexes, in the context of cells and tissues, metabolites, drugs and diseases. I decided to perform a functional and network clustering analysis using the Ingenuity Pathway Analysis (IPA) to understand the pathways that MBD4 might impact on, as well as to identify the potential 'hub' protein candidates for further investigation. I recognise the cohort of the false positives existing in my candidate list may perturb the accuracy of the IPA analysis, however, these annotation results are still capable of targeting to the functional role of MBD4 where carefully considered with the previously suggested phenotype and functions of MBD4 (I will discuss this in the section below). I combined interatomic data from both screens together with previously reported MBD4 interacting proteins, and then mapped them into the IPA protein-protein interaction database, to identify significant biological functions and canonical pathways that MBD4 may involve in (Ingenuity Systems, www.ingenuity.com). The function annotations and network clustering are shown in Figure. 3.9 and Supple Figure. 3.1, which suggest that MBD4 may impact on a number of significant pathways.

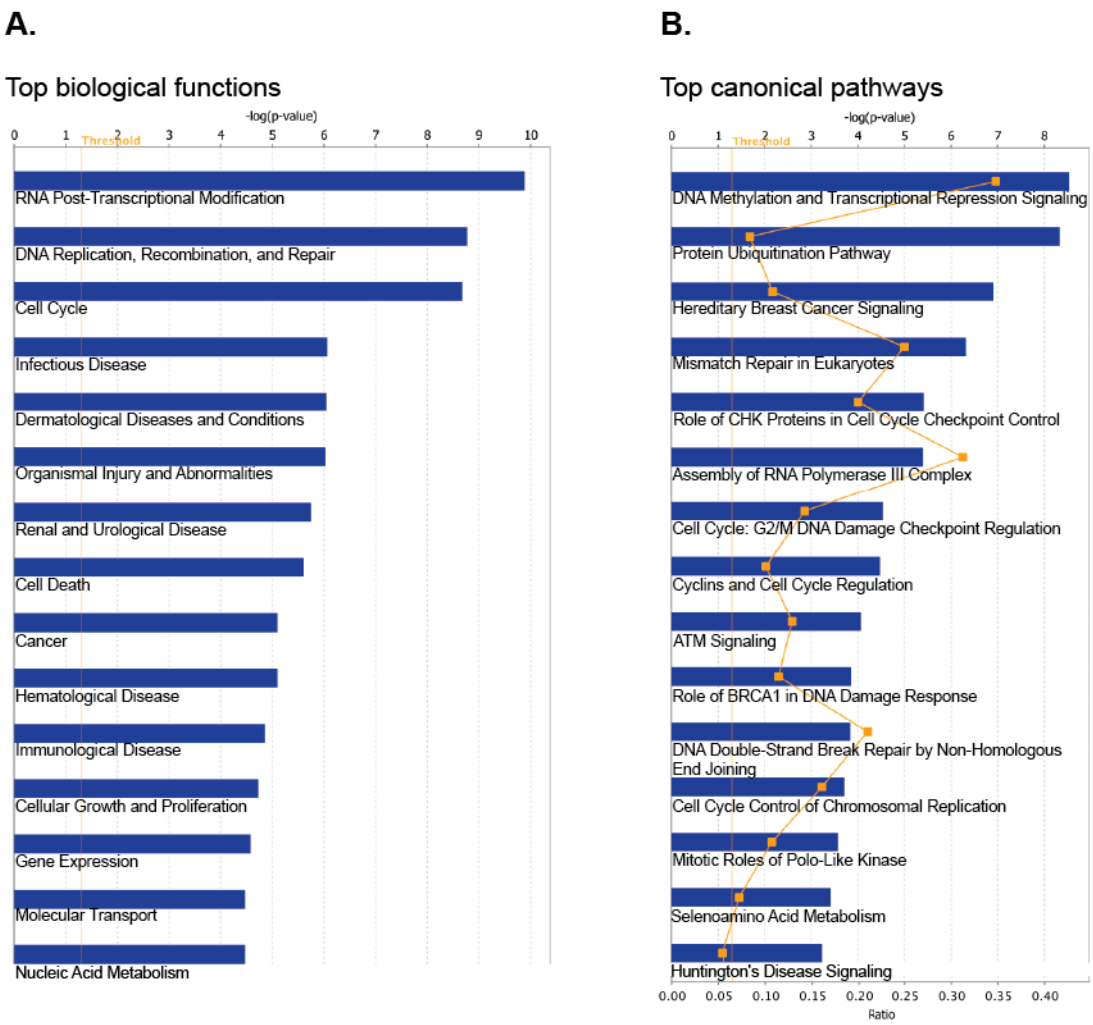


Figure. 3.9. Functional annotations of MBD4 interacting candidates from Y2H and AP/MS screens. All the candidates from both screens together with known MBD4 partners were mapped into IPA databases. Bar charts represent the p-value significance, in which Fisher's exact test was used to calculate a p - value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. (the bigger number at the –log(p-value) scale above represents the more significant category group in the cluster, a threshold line was automatically calculated by the software labelled by the yellow vertical line) The yellow dots in (B) represent the ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway whose scale is at below.

3.3. Discussion

In this chapter I have reported the identification of ~380 potential interacting protein candidates of MBD4 by complementary approaches of Y2H and AP/MS. The top candidates

in either screen and the overlapping candidates from both screens can be thought as MBD4 candidate partners for further investigations. In addition, functional analysis of these protein candidates imply that MBD4 may impact several transcriptional and epigenetic networks along with a number of nuclear pathways that include DNA repair, histone modification and chromatin remodeling, and RNA processing.

MBD4 has been suggested to be involved in many cellular function including DNA repair, apoptosis, DNA demethylation and transcriptional repression. However, the precise molecular function of MBD4 in these processes remains unknown. My screens provide protein candidates for further investigation. Regarding its role in DNA replication, recombination and repair, a group of protein candidates are very interesting (Figure. 3.9 A). These can be mainly divided into two significant protein clusters (Figure. 3.10. A&B); the repair protein cluster (Figure. 3.10 A), and the histone modification and chromatin remodeling protein cluster (Figure. 3.10. B). Particularly, the link between the latter cluster of candidates with MBD4 is largely unknown. Thus it will be very interesting to look into the 'hub' candidates in the context of histone modification and chromatin remodeling, to investigate if MBD4 has a role in recruiting co-regulators and altering the chromosomal organisations to form a suitable local micro-environment for the efficient repair, or independent from its repair function (Figure. 3.9 B). This led to our decision to choose a number of protein candidates within this cluster including PRMT5, UHRF1, and TRIM28 for the further validations in chapter 4. Moreover, I tested two independent methods of Y2H and AP/MS for detection of Mbd4 associations, in addition to the well-designed controls and baits such as MBD domain and full length of MBD4 in Y2H. These enabled me to choose the most reliable candidates from these overlapping lists for further validations (See table 4.1). For example the candidates DEK, SETD1a and USP7, which were identified in both Y2H and AP/MS, and SRSF2 which was the top candidate from both MBD domain and full length screen in Y2H.

It is noteworthy that there may be a number of false positives within my screen candidates. This may be due to the technical limitations of the Y2H, as well as the less stringent threshold I chose in my AP/MS. However, there are three reasons I decided to use this set of non-validated protein candidates to perform the IPA pathway analysis for prediction of MBD4 functions. Firstly, although the overlapping candidates from both screens may be thought as most robust ones for the further investigation, inevitably they lose the clues regarding the connections of the individual MBD4 functions, considering the pathways MBD4 may involved in are functionally relevant to each other. Thus a systemic functional analysis using all these MBD4 screening candidates may take account of all these

functional links in the prediction of MBD4 functions. Secondly, in comparison to the protein interaction knowledge bases of the functional annotation programmes, which are unbiased and stochastic to MBD4 function, the datasets from my MBD4 screening indeed largely tend to be MBD4 associated and relevant to MBD4 functions. Thus the probability of identifying true biological processes and functional clusters that MBD4 may be involved in is much higher than that of false discovery. Thirdly, the existing protein-protein interaction datasets is a scale-free network by and large. This is partially because in the past research, 'valuable' proteins were investigated more extensively and are more well-understood than the others. In such context, the 'hub' candidates that stood out from the clustering analysis would more likely link MBD4 function to well studied pathways that may have significant biological relevance. On the other hand, the existing protein interaction datasets from the knowledge base of IPA is derived from the expert-curated data from the publications and/or verification of protein interaction databases, which may have less false positive data than those of predicted protein interaction databases and possibly GO term databases. A functional annotation and clustering using this IPA datasets could increase the chance to avoid the false positives that is part of the automatic analysis. However, the top 'hub' candidates need to be carefully validated (See Chapter 4) and the canonical pathways and biological functions suggested require thorough consideration together with additional analysis of the misregulated gene expression of *Mbd4*$^{-/-}$ cells and in specific tissue context(s), biological process(es), and/or pathway(s).

Chapter 4: Validation of MBD4 protein interactions and characterisation of the interacting domains of MBD4 responsible for the associations

4.1. Introduction

The protein associations of MBD4 are of interest as through them MBD4 may be involved in a number of pathways and/or they may regulate the activity of MBD4 in biological processes. A number of candidates from screens were selected for validation, and investigations of the interactions between MBD4 and the validated proteins are presented in this chapter.

4.2. Results

4.2.1. Validation of MBD4 Protein associations by Reciprocal co-IP

In Chapter 3, I have identified a large number of candidates that potentially associate with MBD4. To confirm these associations I performed reciprocal coimmunoprecipitations (co-IP) followed by western blotting. Since it was impossible for me to verify the interactions for all of the ~380 MBD4-associated candidates, I chose four candidates from my Y2H Screen and seven candidates from my AP/MS screens (of which three candidate proteins were overlapping the ones from Y2H list) (Table 4.1 & 4.2). Most of these chosen candidates have known functions that may be linked to transcriptional regulation of MBD4 function, including histone modification, chromatin remodelling and components of known repressor complexes. As I mentioned in discussion of chapter 3, a transcriptional repression function for MBD4 has been reported but the precise mechanism is largely unknown. In addition, my top candidate SRSF2 from Y2H might implicate for MBD4 in splicing, thus was also chosen because of its possible biological implications. I did not pick all the candidates that are overlapping between the screens of Y2H and AP/MS because many of their antibodies and plasmids are not immediately available. In addition, I also included three other candidates that were not in my screening lists but are the components of multi-protein complex of chosen candidates, or within the same domain group of my chosen candidates, including CFP1 (form complex with SETD1a), PRMT1 and PRMT3 (same domain group of PRMT5), because they are available when we requested plasmids for the potential candidates. And I thought they were also very good control in parallel of my validations, to test if the whole protein complex is associated with MBD4 or just the single components, or if the whole family proteins sharing significant domain structure interacts with MBD4 or just the protein identified in the screen. I chose MLH1 as my positive control, as it was the most robust

interacting protein of MBD4 in previous publications (Bellacosa et al., 1999, Ruzov et al., 2009). To serve as negative control, I used empty vector and LEDGF/p75, which was in the same vector that my MBD4 was cloned into, and is not expected to associate with MBD4 (Madapura et al., 2012). I chose to use overexpressed tagged proteins because there were no good antibodies for MBD4 at the time. Candidate proteins (tagged with a peptide sequence FLAG-, GFP-, or M-Cherry-) were co-transfected with tagged-MBD4 into HEK 293T cells, and were immunoprecipitated from whole lysates, followed by western blotting (WB) with an MBD4-tag-specific or candidate-tag-specific antibody. A summary of all the validated candidates can be found in Table 4.1.

	Candidates	Y2H	IP Mass Spec	Within same group of or same complex of screen candidates	IP validation	Reciprocol IP
Positive candidates confirmed by reciprocal co-IP	UHRF1	-	Y		Y	Y
	PRMT5	-	Y		Y	Y
	PRMT3	-	-	Y	Y	Y
	PRMT1	-	-	Y	Y	Y
	KAP1	-	Y		Y	Y
Positive candidates confirmed by flag-or GFP-co-IP	USP7	Y	Y		Y	-
	SETD1A	Y	Y		Y	N
	CFP1	-	-	Y	Y	N
Candidates require further validation	ACF1	-	Y		Y	-
	DEK	Y	Y		-	-
Negative candidate	SC35	Y	-	Y	N	N
Negative control	LEDGF/P75	-	-	-	N	-
Positive control	MLH1	-	-	Y	Y	-

Table 4.1. List of MBD4 associated protein candidates for the further validations by reciprocal co-IP.

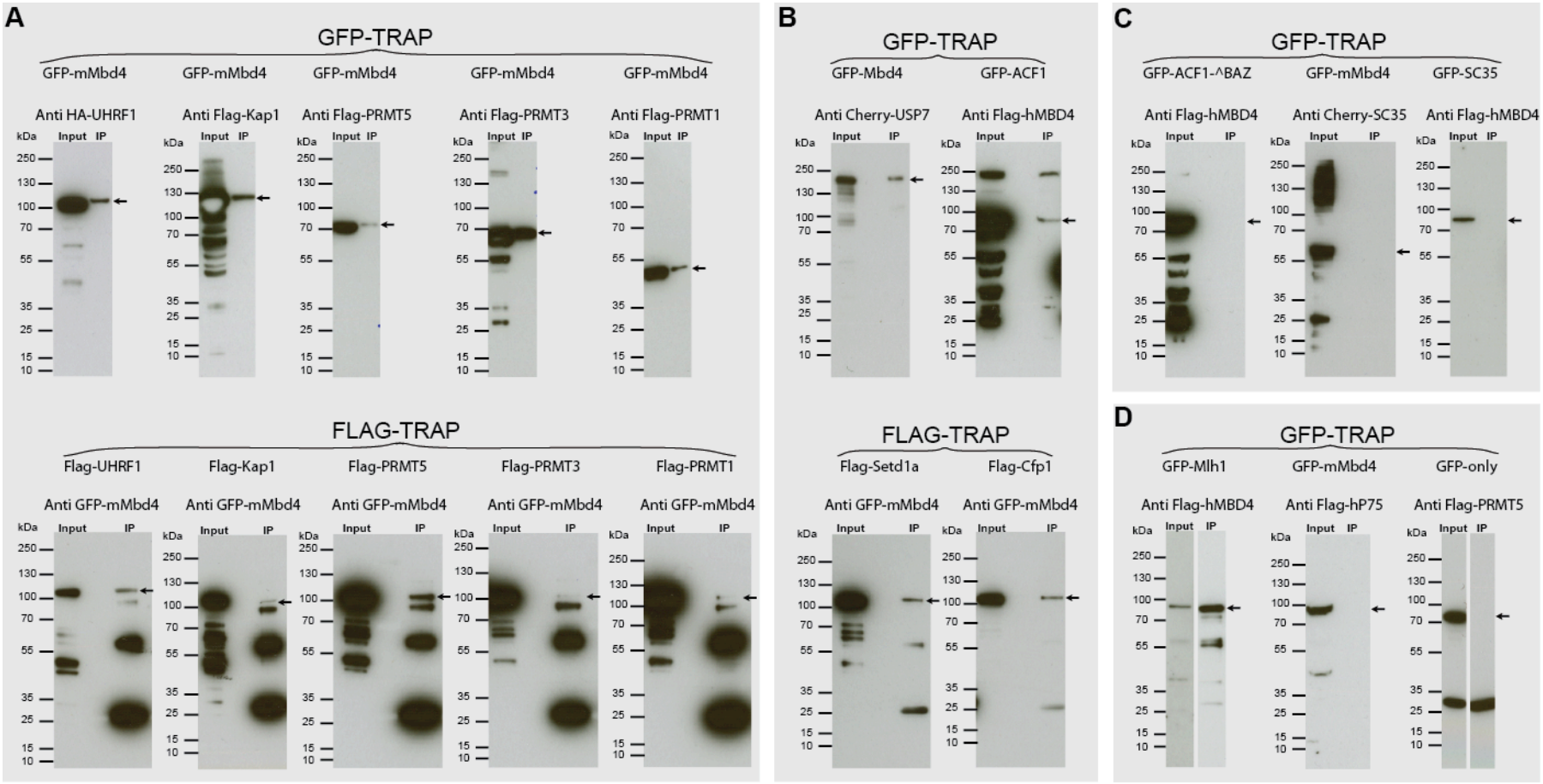

A
GFP-TRAP
GFP-mMbd4
GFP-mMbd4
GFP-mMbd4
GFP-mMbd4
GFP-mMbd4
Anti HA-UHRF1
Anti Flag-Kap1
Anti Flag-PRMT5
Anti Flag-PRMT3
Anti Flag-PRMT1
FLAG-TRAP
Flag-UHRF1
Flag-Kap1
Flag-PRMT5
Flag-PRMT3
Flag-PRMT1
Anti GFP-mMbd4
B
GFP-TRAP
GFP-Mbd4
GFP-ACF1
Anti Cherry-USP7
Anti Flag-hMBD4
FLAG-TRAP
Flag-Setd1a
Flag-Cfp1
C
GFP-TRAP
GFP-ACF1-^BAZ
GFP-mMbd4
GFP-SC35
Anti Flag-hMBD4
Anti Cherry-SC35
D
GFP-TRAP
GFP-Mlh1
GFP-mMbd4
GFP-only
Anti Flag-hMBD4
Anti Flag-hP75
Anti Flag-PRMT5
Input
IP
kDa

Figure. 4.1. Confirmation of MBD4 protein associations by reciprocal co-IP. HEK293T cells expressing FLAG- or GFP- tagged MBD4 and a given candidate interactor, expressed exogenously as an HA-, FLAG-, GFP-, or Cherry- tagged protein, were lysed and their total cell extracts were IPed with an antibody directed at the candidate or tagged MBD4. Associations between MBD4 and the candidates were analysed by sequential Western blotting using indicated antibodies. (A) 5 candidates UHRF1, KAP1, PRMT5, PRMT3, and PRMT1 were confirmed by reciprocal co-IP. (B) 4 candidates USP7, ACF1, SETD1a, and CFP1 were confirmed by one FLAG- or GFP- co-IP. (C) SRSF2(SC35) validation was negative. (D) MLH1 was positive control, and LEDGF/P75 and GFP only were negative controls.

The positive control MLH1 showed a robust interaction with MBD4, and the negative control LEDGF/P75 were very clear negative in the IP lane, suggesting that my reciprocal co-IPs were technically reliable (Figure. 4.1 D). Nine out of eleven protein candidates tested showed positive associations with MBD4 by reciprocal co-IPs or single side co-IP (Figure. 4.1 A-B), suggesting that these proteins are true Mbd4-interactors (either direct or indirect). Among these nine validated protein candidates, five proteins including UHRF1, PRMT5, PRMT3, PRMT1, and KAP1 were validated by reciprocal co-IP. Among the four other proteins validated by single side co-IPs, USP7 showed robust binding with MBD4 in two independent co-IPs where MBD4 was tagged by FLAG or GFP. Another three protein candidates including SETD1a, CFP1, and ACF1 can be IPed from GFP- or FLAG-tagged MBD4, but failed to be detected in reciprocal IP. However, CFP1 did indeed affect the expression level and pattern of MBD4, which is shown later in this chapter and ACF1 did tightly colocalise with Mbd4 within nuclei, also shown later in this chapter, indicating their associations with MBD4 are true positive. In contrast, two protein candidates DEK and SRSF2/SC35 were negative in my co-IPs (Figure. 4.1 C). However, I noticed the FLAG-DEK expression is pretty low in the input lane, and the GFP-DEK was undetectable under the fluorescent microscopy, suggesting it needs further investigation. In addition, SRSF2/SC35 was originally screened out from my Y2H, and was the top hit candidate for both full length and MBD domain Y2H screens. Due to the technical diversity between Y2H and IP, which I have discussed in Chapter 3, it is not unusual that the top hit candidate SC35 could not be detected in IP validation, as the interaction between SRSF2/SC35 and MBD4 may be transient or very weak. Notably SRSF2/SC35 did indeed inhibit the expression level of MBD4, which is shown in Chapter 4.2.3, indicating there is certain connection between SRSF2/SC35 and MBD4. All the co-IPs were repeated at least three times by different wash buffers, wash times, and with/without DNase, except ACF1, which was positive at the first time but negative at the second essay. However, its mutant GFP-ACF-^BAZ was negative (Figure. 4.1. C left), in parallel with its positive result (Figure. 4.1. B upper right), suggesting

ACF1 is a true positive MBD4 interacting candiate but further validation is required. In sum, the validations by reciprocal IP showed the overall robustness of my Y2H and IP/MS screens, and identified a cohort of significant MBD4 interacting proteins for further investigations.

4.2.2. Subcellular localisations of MBD4-Interacting Protein candidates with co-transfection of MBD4 analysed by immunofluorescence microscopy

To test if the chosen candidates occupy the same cellular space with MBD4 (in 2 dimensions), I performed co-transfection followed by immunofluorescent (IF) microscopy to determine their potential co-localisations and subcellular distributions in mouse CMT93 and human HEK293T cell lines. If the proteins are concentrated in the same cellular space, it would support the idea that they may form stable complexes with MBD4. Immunofluorescent microscopy was performed on fixed mouse CMT93 cells after transient transfection of FLAG- or GFP- tagged MBD4 and one of the chosen MBD4 interacting candidates. The MBD4 and its interacting candidates were either fused to GFP-, M-Cherry-fluorescent peptides, or FLAG- tagged which was detected by antibody against FLAG peptide and Texas Red- conjugated secondary antibody. A summary of IF microscopy for all the validated candidates can be found in Table 4.2.

Candidate	Nuclear expression?	Cytoplasmic expression?	Colocalisation with MBD4?	Relocated after co-transfection with MBD4?	Chromatin changes when co-transfection with MBD4?
UHRF1	Y	N	Y		Y
PRMT5	Y /several spots	Y	N		Y
PRMT3	Y	Y	Y		
PRMT1	Y	Y	Y /partially		
KAP1	Y	Y	N		
USP7	Y	Y	Y	Y	Y
SETD1A	Y	Y /occasionally	N		Y
CFP1	Y	Y	Y /partially		
ACF1	Y	N	Y		Y
DEK	Y	N	Y /partially		
SC35	Y	Y /occasionally	Y /partially		
MLH1	Y	Y	Y	Y	Y

Table 4.2. List of MBD4 associated protein candidates for IF co-localisation.

MBD4 and MLH1

Overexpressed MBD4 protein, fused with either GFP or Flag tag, is localized to the nucleus, and exhibits concentrated signal at DAPI bright spots, resembling the endogenous MBD4 distribution (Figure. 4.2) (Hendrich and Bird, 1998). The DAPI bright spots correspond to methylated satellite DNA, a natural ligand for MBD4 (Ruzov et al., 2009). Non-heterochromatic staining of GFP or Flag tagged MBD4 is also clearly observed in CMT93 cells. Transfection of MBD4 using vectors expressing GFP-, or FLAG- tagged full-length mouse or human MBD4 results in the similar nucleus distribution (Figure. 4.2), except occasionally cytoplasmic expression can be observed for FLAG tagged human MBD4 in CMT93 cells (Figure. 4.2 C). The bright staining of MBD4 was concentrated in areas of DAPI bright spots, consistent with the localization of MBD4 to heterochromatic regions of the nucleus (Ruzov et al., 2009). Co-localization experiments were conducted to characterize the associations between the nucleus distributions of MBD4 and its interacting candidates. The mouse CMT93 Cell line is a colon cancer cell line with intact heterochromatin sites, and was chosen to perform the co-transfection with vectors expressing epitope- tagged (GFP or FLAG) full length Mbd4 with epitope-tagged (GFP, M-Cherry, or FLAG) interacting candidate proteins. The human HEK293T cell line was also used (see section 4.2.3).

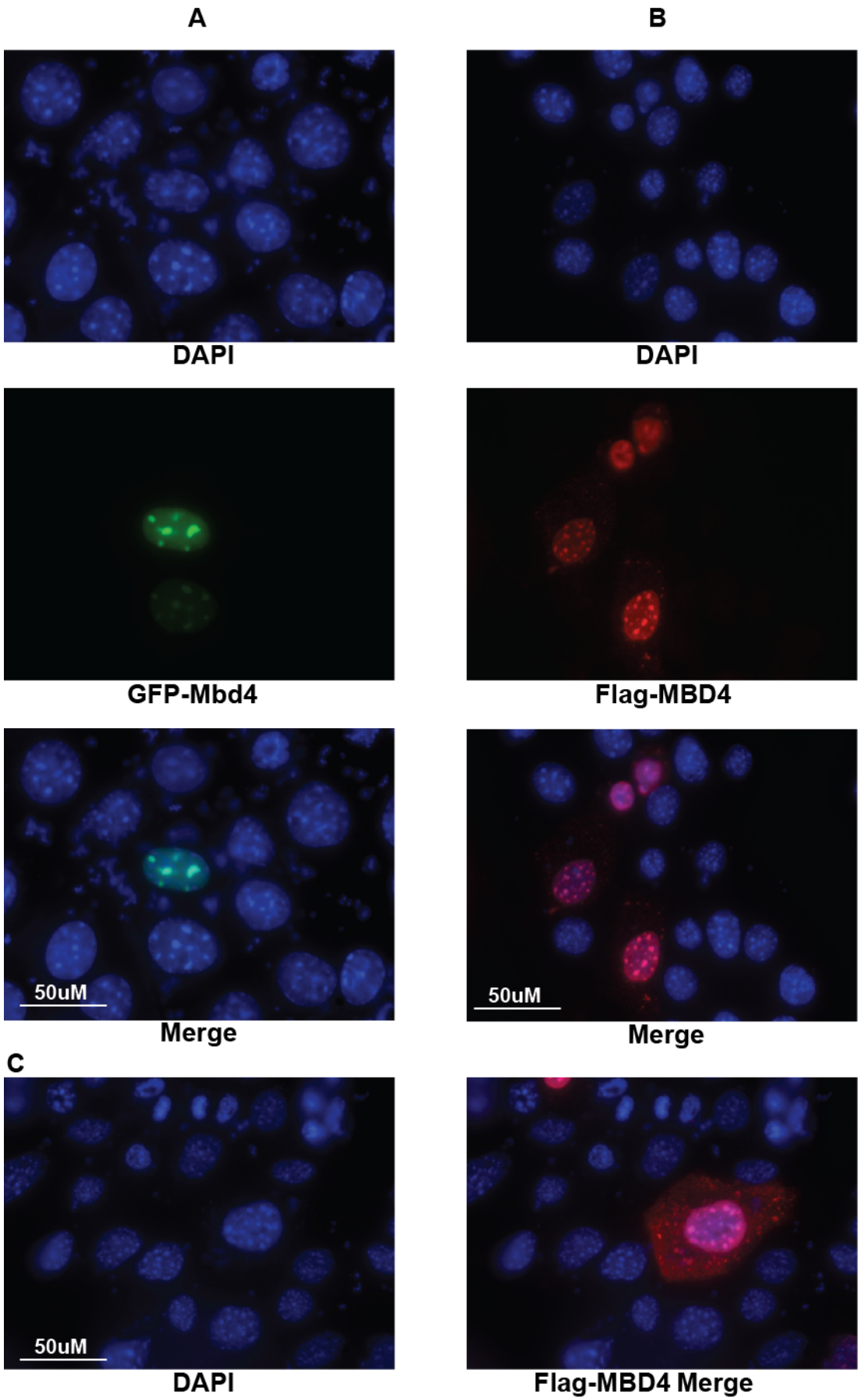

A
B
DAPI
DAPI
GFP-Mbd4
Flag-MBD4
50uM
50uM
Merge
Merge
C
50uM
DAPI
Flag-MBD4 Merge

Figure. 4.2. MBD4 localises to bright DAPI spots that are tightly associated with heterochromatic condensed chromatins. A. Nuclear distribution of bright spots containing GFP-mMBD4 was compared with DAPI staining in mouse CMT93 cells. B. Nucleus distribution of bright spots containing FLAG-hMBD4 was compared with DAPI staining in CMT93 cells. Vectors expressing FLAG tagged human MBD4 fusion proteins were transiently transfected into mouse CMT93 cells, and were detected using anti-FLAG antibody and Texas Red- conjugated secondary antibody. Fixed cells were counterstained with DAPI. The nucleus distribution of overexpressed Mbd4 resembles that of endogenous mMBD4 (Hendrich and Bird, 1998). All the cells successfully transfected vectors containing GFP-mMBD4 and almost all the FLAG-hMBD4 show this pattern. C. Occasionally cytoplasmic expression of FLAG tagged hMBD4 can be observed, but the pattern of their nucleus expression is the same as the pattern described above.

I used MLH1 protein as my positive control. In CMT93 cells, the single transfection of GFP-Mlh1 showed a diffused localisation pattern in the nucleus and some staining in the cytoplasm, and the DAPI bright spots were largely excluded from the distribution of MLH1, fitting with previously published results for MLH1 locolisation (Figure. 4.3 A) (Ruzov et al., 2009). In contrast, after co-transfection of GFP-MLH1 with FLAG-hMBD4, GFP-MLH1 exhibited a distinct distribution pattern. The distribution of MLH1 became condensed, and concentrated at DAPI bright spots, exactly colocalising with MBD4 bright dots in the nucleus, where was previously reported to be associated with heterochromatic chromatins (Figure. 4.3 B) (Ruzov et al., 2009). My IF results of the positive control MLH1 confirmed our previous finding that overexpressed MBD4 recruits MLH1 to DAPI bright heterochromatic sites in *p53*$^{-/-}$ cells (Ruzov et al., 2009), indicating my protocol for IF works well. In addition, the successful reproduction of the recruitment of MLH1 to heterochromatic chromatin by MBD4 in our new cell model of CMT93 cells served a excellent positive control for my IF protocol and cell model in my following observations.

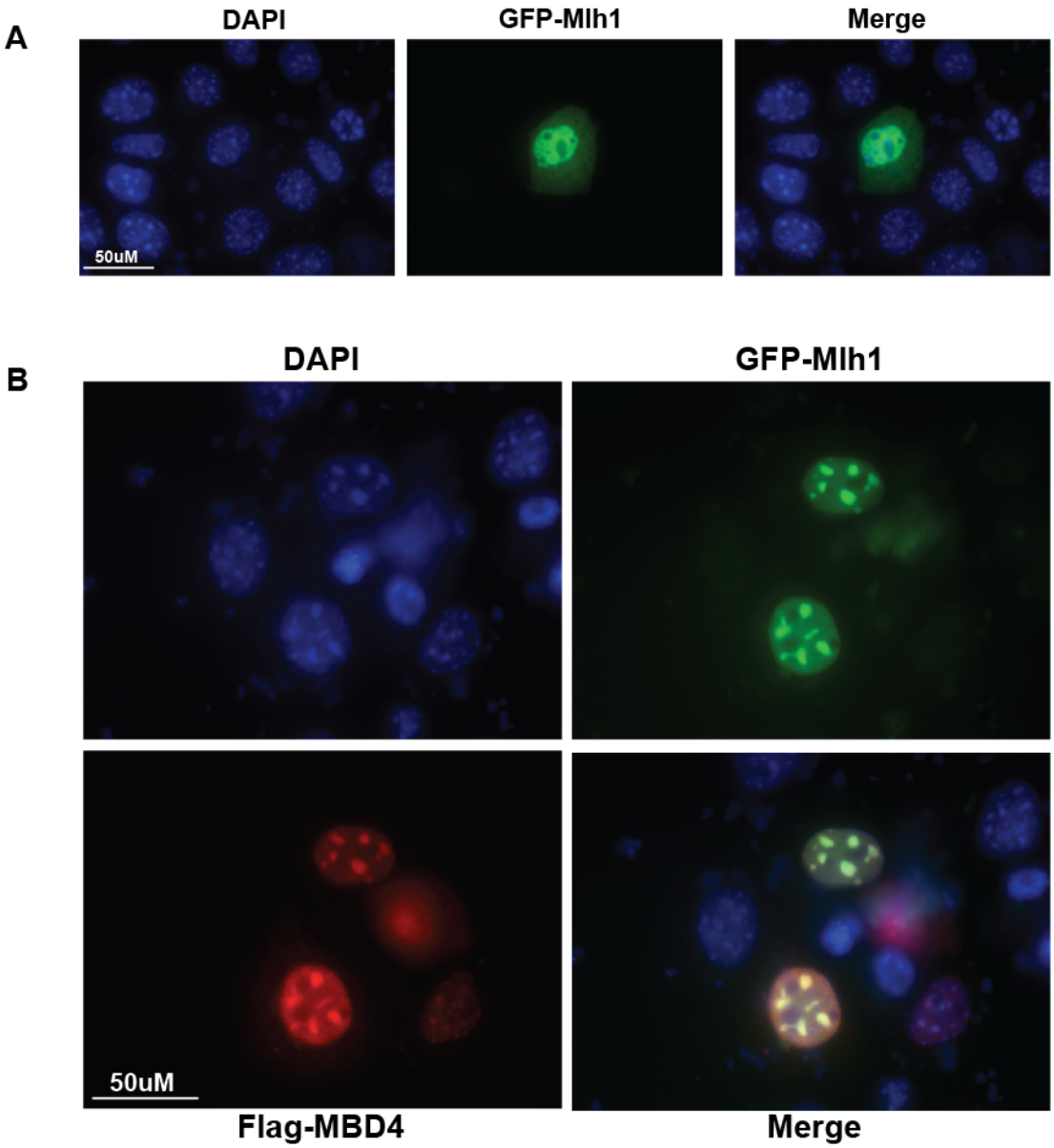


Figure. 4.3. MBD4 recruits MLH1 to bright DAPI spots that are tightly associated with heterochromatic condensed chromatins. A. Diffused nucleus distribution pattern of MLH1 by itself where heavily stained DAPI spots were excluded from the MLH1 distribution. Vectors containing GFP-mMLH1 were transiently transfected, and compared with DAPI staining in CMT93 cells. B. Co-transfection of FLAG-hMBD4 and GFP-MLH1 into mouse CMT93 cells results in the recruitment of GFP-MLH1 to DAPI bright spots. Vectors expressing FLAG-hMBD4 and GFP-MLH1 were transiently transfected into mouse CMT93 cells, and were detected using anti-FLAG antibody and Texas Red- conjugated secondary antibody. These results repeated our previous findings in (Ruzov et al., 2009) and served as positive control. All the cells single- or double transfected show this pattern (See supplementary Figures 4.3_4.4 A1&A2).

USP7, UHRF1, and ACF1

My MBD4 interacting protein candidate USP7 is M-Cherry tagged and exhibits a nuclear distribution (and some in the cytoplasm) that is similar to that of the positive control MLH1 in both single transfection and co-transfection with MBD4. M-Cherry-USP7 by itself distributed a diffused pattern in the nucleus (Figure. 4.4 A), and when cotransfected with MBD4, it is clear that MBD4 recruits USP7 to DAPI bright spots (Figure. 4.4 B). This supports the idea that these proteins can interact.

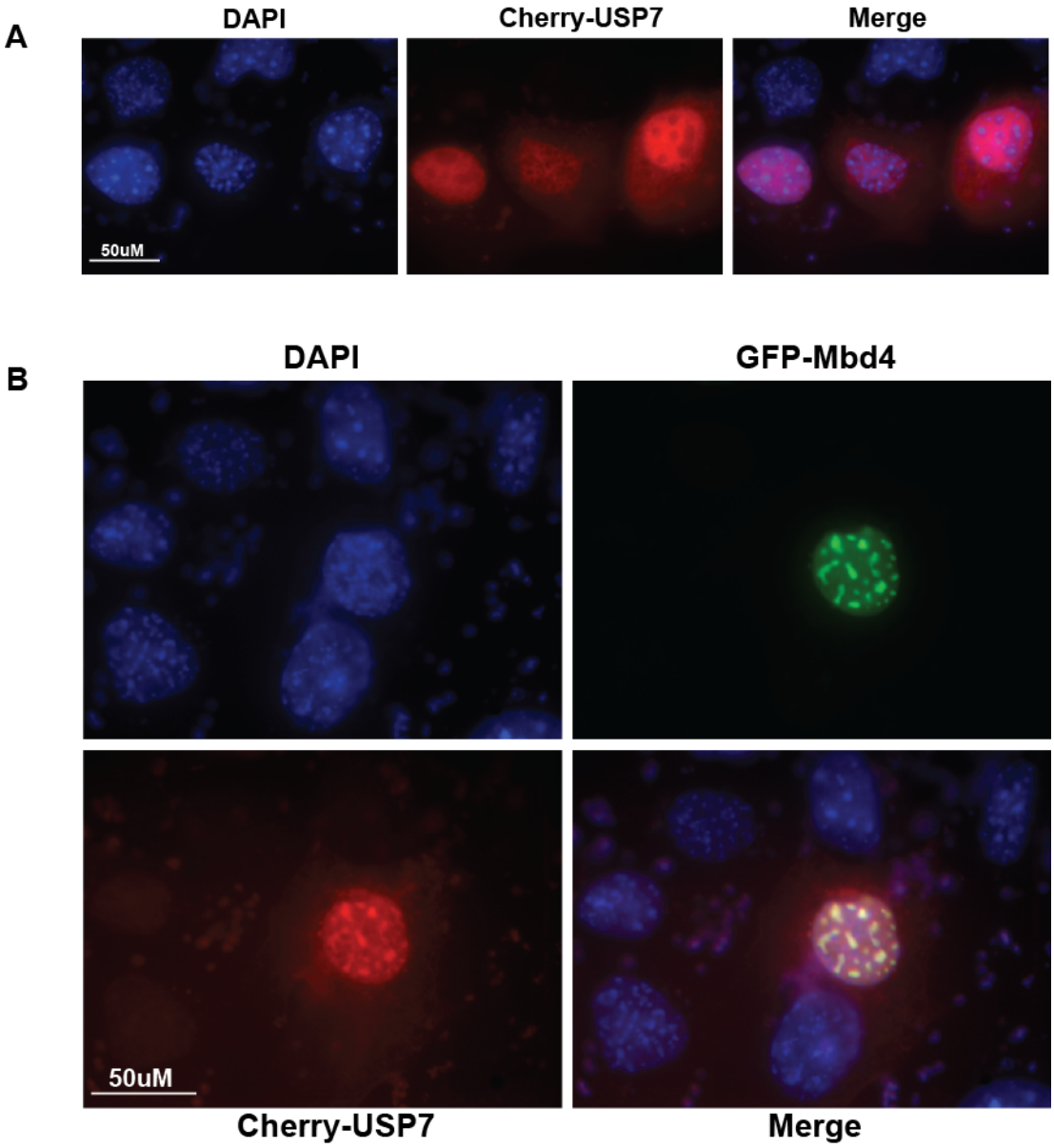
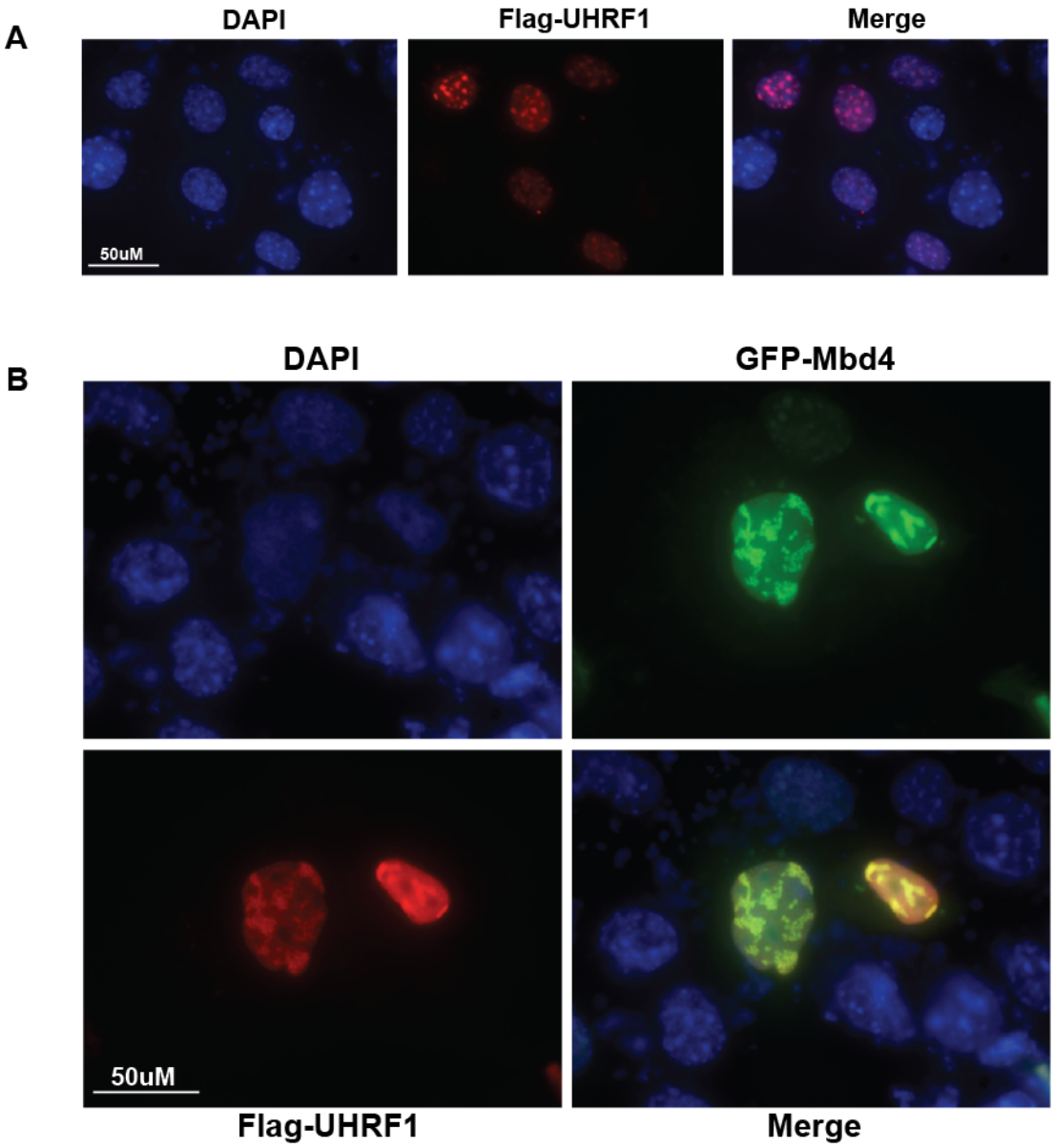


Figure. 4.4. MBD4 recruits USP7 to bright DAPI Spots that are tightly associated with condensed heterochromatic chromatins. A. Diffused nucleus distribution pattern of USP7 by itself where heavily stained DAPI spots were excluded from the USP7 distribution. Vectors containing Mcherry-USP7 were transiently transfected, and compared with DAPI staining in CMT93 cells. B. Co-transfection of GFP-mMBD4 and Mcherry-USP7 into mouse CMT93 cells results in the recruitment of Mcherry-USP7 to DAPI bright spots. Vectors expressing GFP-mMBD4 and Cherry-USP7 were transiently transfected into mouse CMT93 cells. All the cells single- or double transfected show this pattern (See supplementary Figures 4.3_4.4 B1&B2).

The other MBD4 interacting protein candidate UHRF1 exhibited a nuclear distribution that was condensed and colocalised with DAPI bright dots when FLAG-UHRF1 was overexpressed in CMT93 cells. This is similar to the subcellular distribution of the endogenous UHRF1 protein (Bostick et al., 2007) (Figure. 4.5 A). In the cells with a single transfection of FLAG-UHRF1, occasionally (one of seven in supplementary figure C1) some minor changes in the chromatin morphology is observed where UHRF1 accumulates can be observed in the nucleus. This is associated with a reduced DAPI bright spots but the size of the cells was normal. However, when FLAG-UHRF1 was co-transfected with GFP-MBD4, the chromatin where FLAG-UHRF1 and GFP-MBD4 accumulated exhibited dramatic changes (Figure. 4.5 B&C&D) (four of six in supplementary figure C2), and the DAPI bright dots were drastically reduced (all the double transfected cells). The size of the co-transfected cells was much bigger than the untransfected controls (four of six in supplementary figure C2), and the chromatins were either very condensed in a fragmented manner or a continued distribution. In both cases, MBD4 was tightly colocalised with UHRF1, and some of the DAPI bright staining was drastically reduced (Figure. 4.5 B).


A
DAPI
Flag-UHRF1
Merge
50uM
B
DAPI
GFP-Mbd4
50uM
Flag-UHRF1
Merge

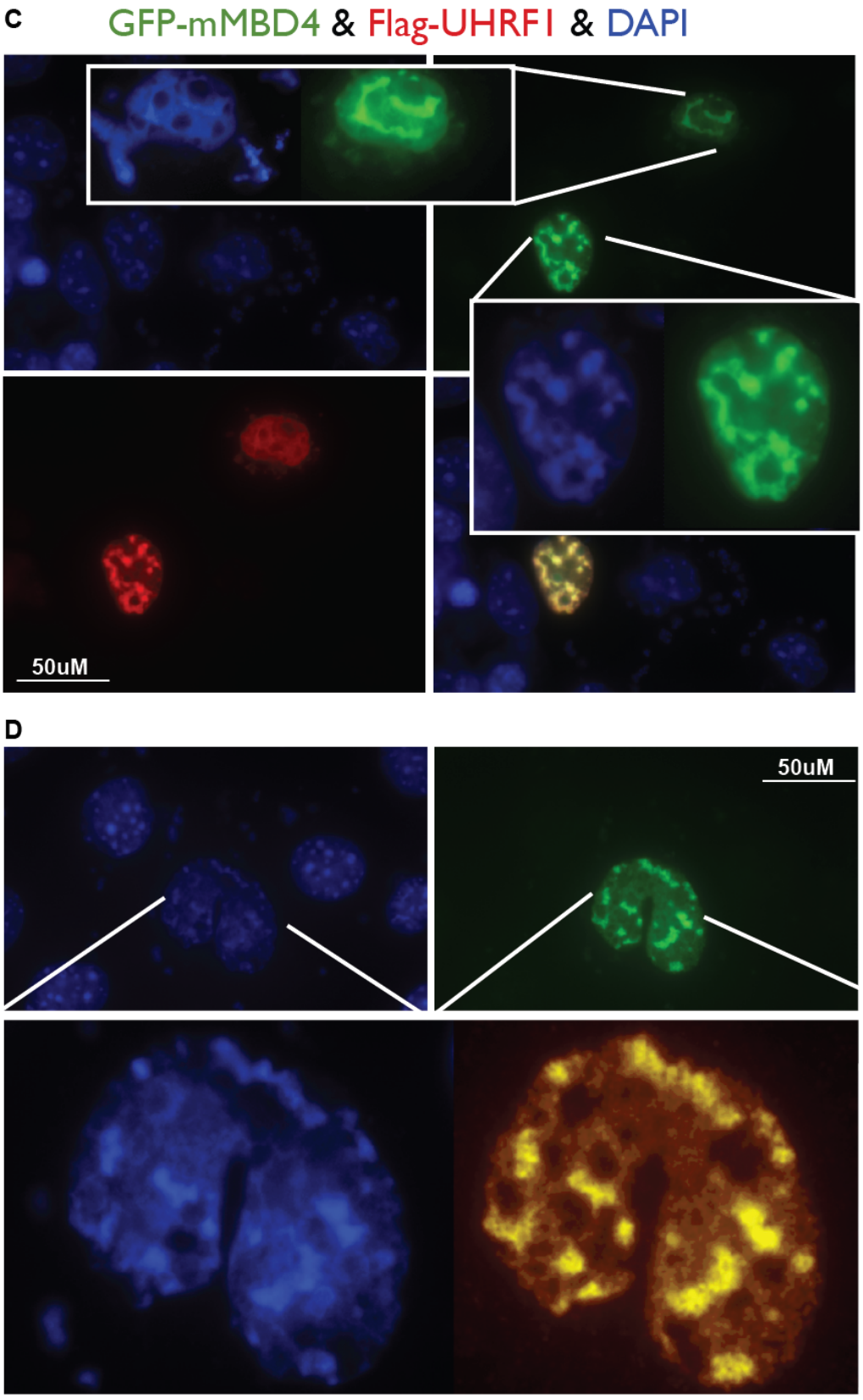

C
GFP-mMBD4 & Flag-UHRF1 & DAPI
50uM
D
50uM

Figure. 4.5. MBD4 colocalises with UHRF1 causing dramatic changes of chromatins & disappearance of DAPI bright spots. A. Nucleus distribution of bright spots containing FLAG-UHRF1 was compared with DAPI staining in CMT93 cells. Vectors expressing FLAG tagged UHRF1 fusion proteins were transiently transfected into mouse CMT93 cells. Fixed cells were counterstained with DAPI. B. Co-transfection of GFP-mMBD4 and FLAG-UHRF1 into mouse CMT93 cells results in the significant changes at chromatic sites (see C&D). Overexpressed MBD4 and UHRF1 tightly colocalise together, and bright DAPI spots associated with heterochromatic chromatins disappeared. In addition, the nucleus size of over half of all the double transfected cells are much bigger than the single transfected cells containing MBD4 or UHRF1 (See C&D, and also see supplementary Figure 4.5_4.6 C1&C2). Vectors expressing GFP-mMBD4 and FLAG-UHRF1 were transiently transfected into mouse CMT93 cells, and were detected using anti-FLAG antibody and Texas Red- conjugated secondary antibody. All the single transfected cells and over half of double transfected cells show this pattern (See supplementary Figure. 4.5_4.6 C1&C2). Other double transfected have no change at their nucleus size but cells also show significant chromatic changes (See supplementary Figure. 4.5_4.6 C1&C2).

A third candidate ACF1 showed a tight co-localisation with MBD4 (Figure. 4.6). ACF1 co-localises with DAPI bright spots in the nucleus when singly overexpressed. In the cells co-transfected with ACF1 and MBD4, they were both concentrated and some of them (~70-80%) became larger condensed bight spots, colocalising with the DAPI bright regions, exhibiting a possible chromatin remodeling event happening at the heterochromatic sites.

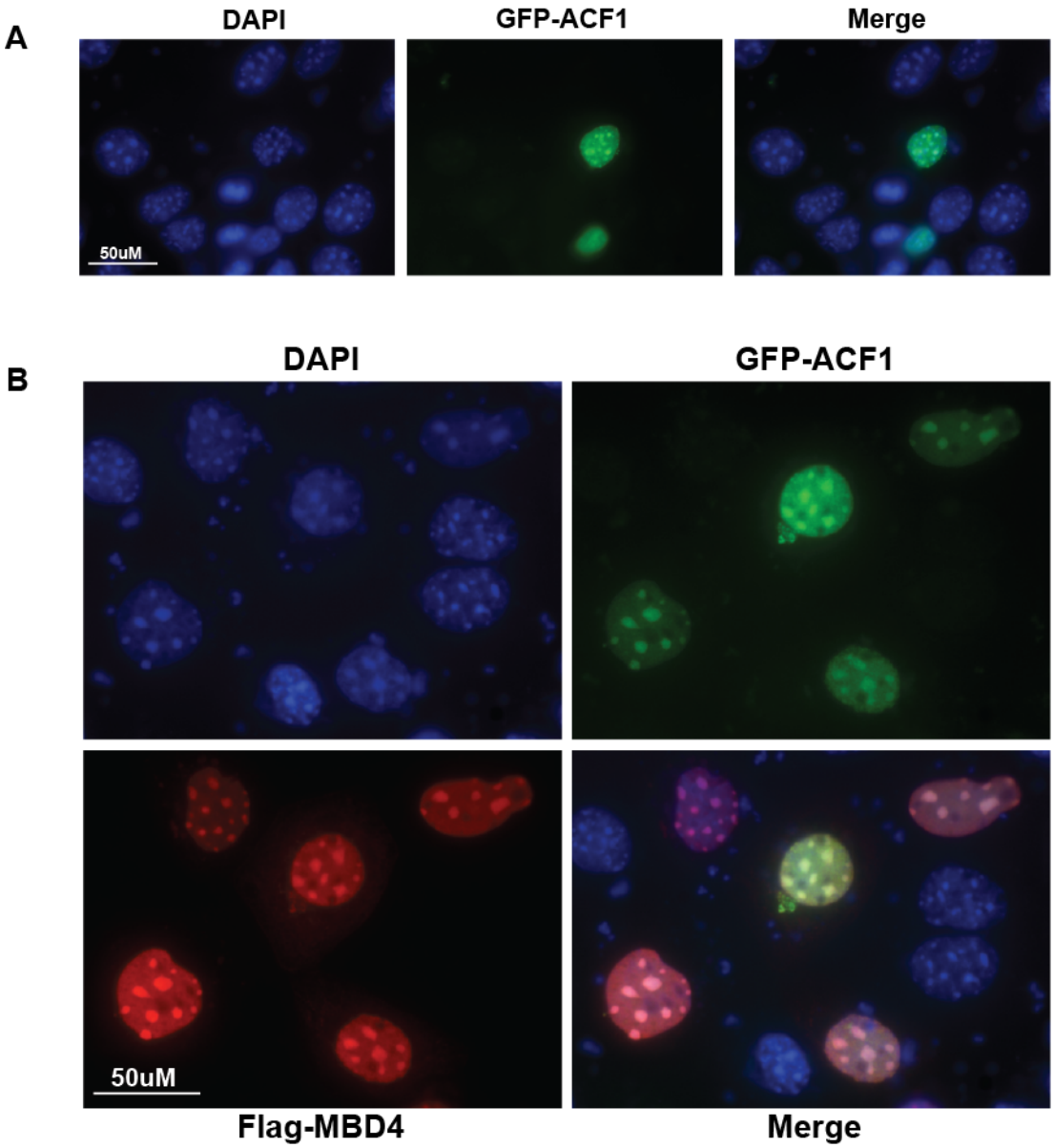
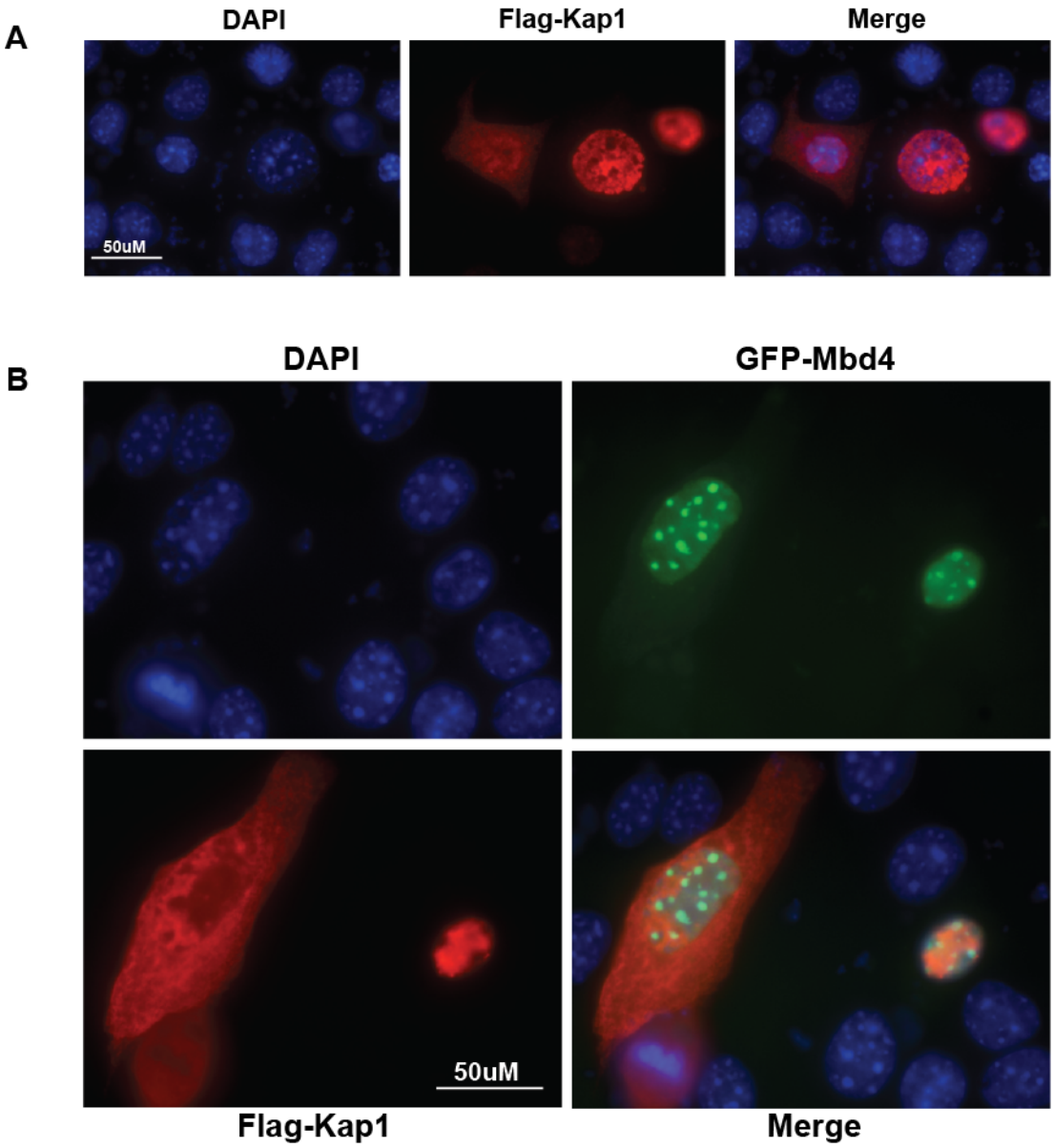


Figure. 4.6. MBD4 colocalises with ACF1 at DAPI bright spots showing a condensed pattern of distribution. A. Nucleus distribution of speckles containing GFP-ACF1 was compared with DAPI staining in CMT93 cells. Vectors expressing GFP tagged ACF1 fusion proteins were transiently transfected into mouse CMT93 cells. Fixed cells were counterstained with DAPI. B. Co-transfection of FLAG-hMBD4 and GFP-ACF1 into mouse CMT93 cells. Overexpressed MBD4 and ACF1 tightly colocalise together in a condensed manner at bright DAPI spots. Vectors expressing FLAG-hMBD4 and GFP-ACF1 were transiently transfected into mouse CMT93 cells, and were detected using anti-FLAG antibody and Texas Red-conjugated secondary antibody. ~70-80% of single and double transfected cells show this pattern (See supplementary Figure. 4.5_4.6 D1&D2).

KAP1, PRMT5, CFP1 and SETD1a

In CMT93 cells, overexpressed FLAG-KAP1 exhibits a general diffused distribution in whole cell or only in the nucleus but it is excluded from the DAPI bright spots (Figure. 4.7 A). This staining pattern of KAP1 can be observed in all the single FLAG-KAP1 transfection and the co-transfection of FLAG-KAP1 with GFP-MBD4. Occasionally several large condensed staining foci of FLAG-KAP1 can be observed in the cytoplasm and tightly attached to the nuclear membrane in both single- and co-transfection (1 of 5 in Supplementary Figures E1&E2). It is noteworthy that the KAP1 distribution was reported to be very dynamic depend on the cell differentiation stage, the cell types and endogenous or ectopic conditions, thus it is not surprising that the distribution of overexpressed KAP1 in CMT93 cells was very different and might be artificial. However, it is notable that KAP1 distribution in all my essays was excluded from GFP-MBD4 and DAPI bright spots, but tightly surrounded them (Figure. 4.7 B&C), which is consistent with the previous observations that KAP1 was not seen concentrated at the visible domains of constitutive heterochromatin in differentiated mouse cells, even in cells where KAP1 is concentrated there (Briers et al., 2009).


A
DAPI
Flag-Kap1
Merge
50uM
B
DAPI
GFP-Mbd4
50uM
Flag-Kap1
Merge

c

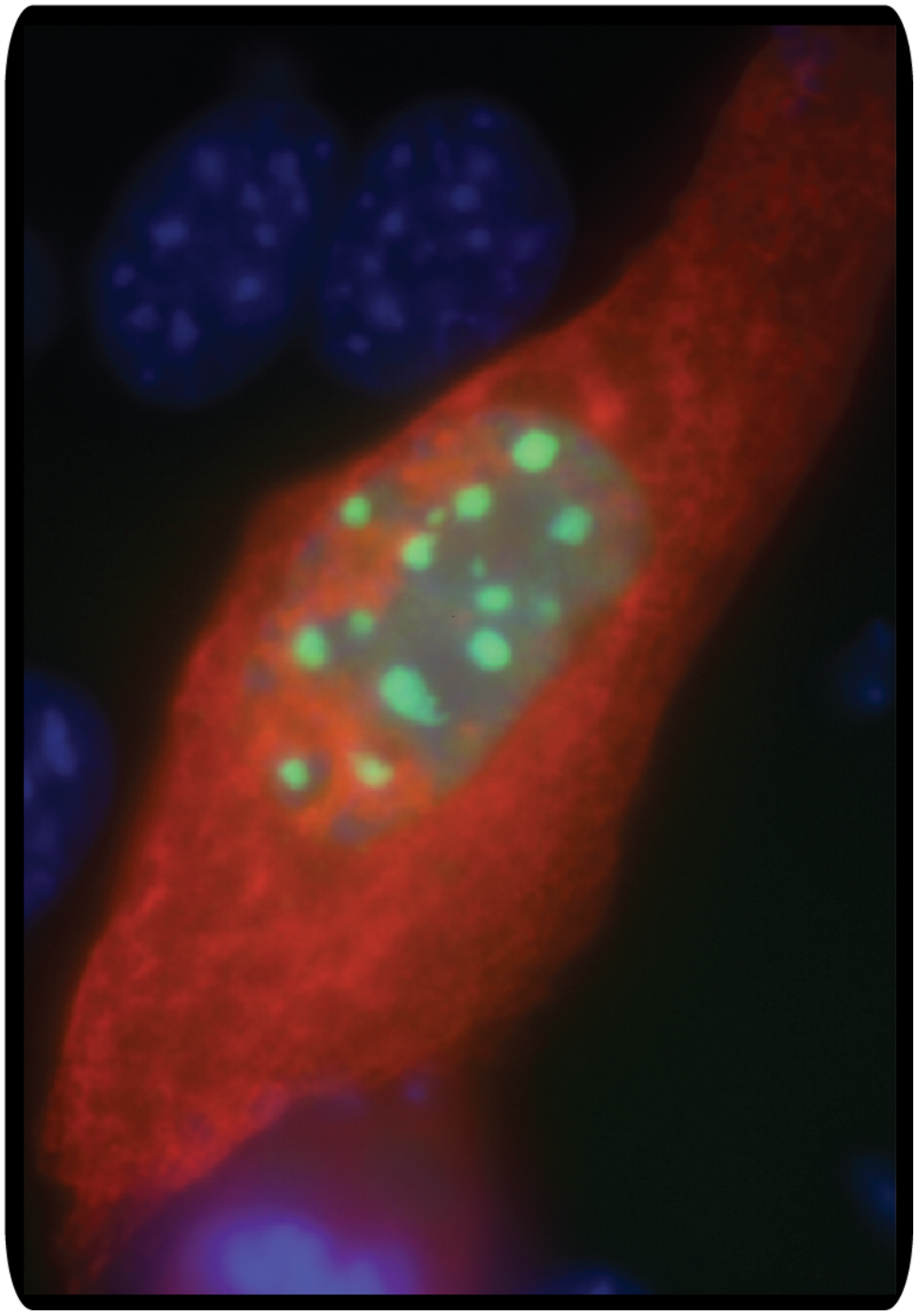
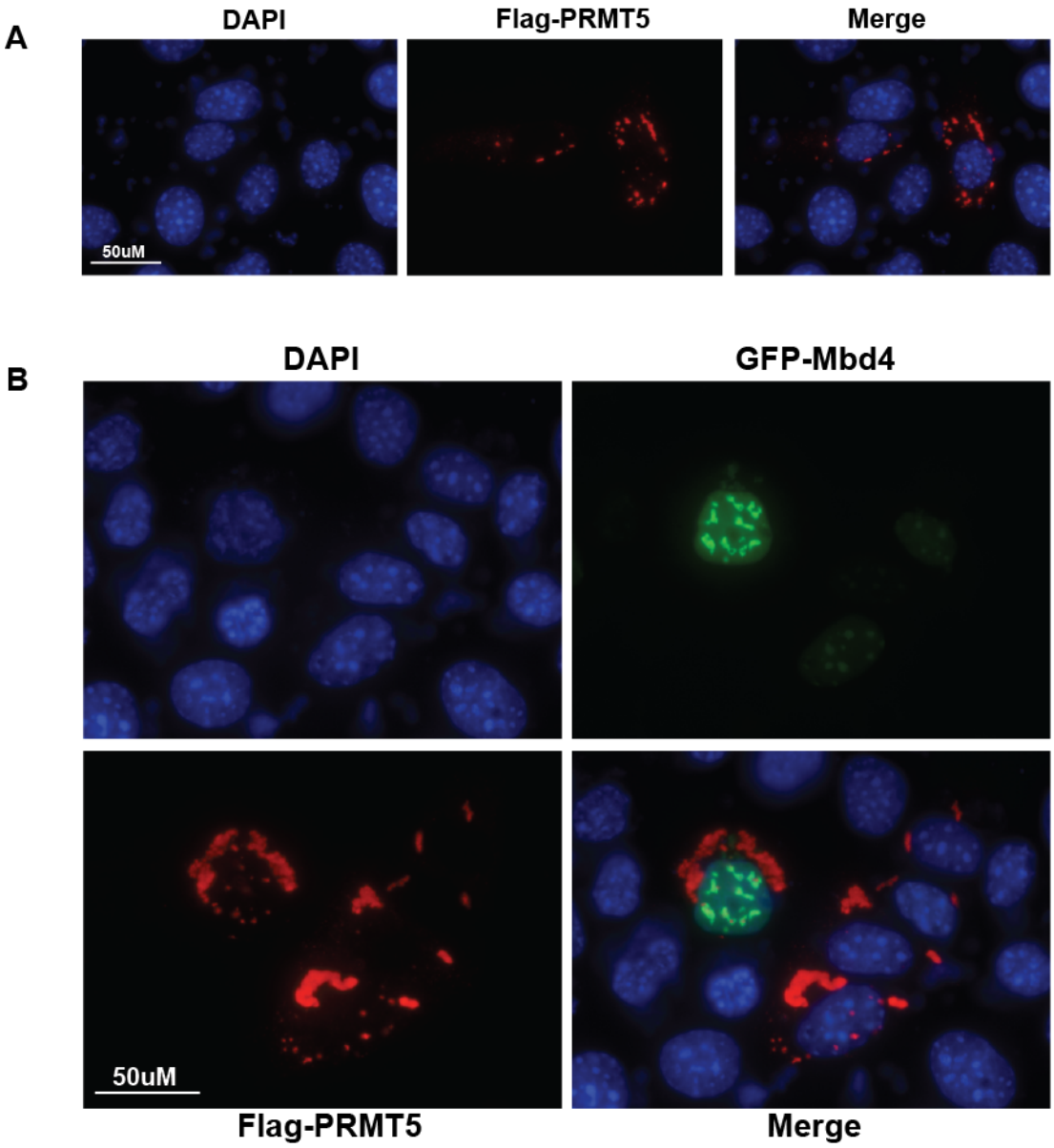


Figure. 4.7. KAP1 does not colocalise with MBD4 and bright DAPI spots but its diffused distribution tightly surrounds Mbd4 and DAPI bright spots. However, it is noteworthy that the unusual distribution of overexpressed KAP1 may be artificial. A. Nucleus distribution of KAP1 was compared with DAPI counterstaining. Vectors containing FLAG-KAP1 were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed KAP1 exhibits diffused general distribution in whole cell or only in nucleus. The bright DAPI spots were excluded from such diffused distribution of KAP1. B. GFP-mMBD4 and FLAG-KAP1 were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots, whereas KAP1 shows diffused general distribution in whole cell or in nucleus the same as controls, but excluded from DAPI and MBD4 bright spots. (See C) All the cells single- or double transfected show this pattern (See supplementary Figure. 4.7_4.8 E1&E2).

Consistent with the previous publication regarding the dynamic distribution of PRMT5 during development and in different cell types (Tee et al., 2010, Zhao et al., 2009), FLAG-PRMT5 in mouse CMT93 cells showed highly condensed spots in cytoplasm, but surrounded the nucleus tightly (Figure. 4.8 A). However, several single or twin bright spots of FLAG-PRMT5 are clearly located within nucleus in both single- and co-transfection with MBD4 (Figure. 4.8 A&B). The bright spots of FLAG-PRMT5 do not colocalise with MBD4 and DAPI bright spots (Figure. 4.8 C), however, in about half of the co-transfected cells, dramatic changes were observed at the condensed chromatins where MBD4 and DAPI bright spots colocalised (See Supple Figure. F1&F2). Notably, in such regions within the nucleus, FLAG-PRMT5 spots were surrounded by MBD4 located chromatin, indicating a potential cross-talk between MBD4 and PRMT5 at the loci (Figure. 4.8 C). In contrast, Overexpressed FLAG tagged PRMT1 and PRMT3 showed a general diffused staining in both cytoplasm and nucleus but some of the bright DAPI spots were excluded (Figure. 4.9-4.10). In both single- and co-transfection with MBD4, FLAG tagged PRMT1 and PRMT3 partially colocalised with MBD4 and DAPI spots, except for the nucleolus (Figure. 4.9-4.10).


A
DAPI
Flag-PRMT5
Merge
50uM
B
DAPI
GFP-Mbd4
50uM
Flag-PRMT5
Merge

c

Figure. 4.8. Ectopic PRMT5 mainly expresses in cytoplasm and does not colocalise with MBD4 and bright DAPI spots. Although a few PRMT5 expressions does distribute as spots in nucleus, they are clearly not colocalised with MBD4 spots. However, half of the double transfected cells show a pattern that the PRMT5 spots in nucleus are surrounded by MBD4 expression, and dramatic changes in chromatin organisation can be observed at MBD4 expression sites (See C & also see supplementary Figure. 4.7_4.8 F1&F2). A. Nucleus distribution of PRMT5 was compared with DAPI counterstaining. Vectors containing FLAG-PRMT5 were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed PRMT5 exhibits condensed distribution in cytoplasm attaching nucleus, and several separated PRMT5 spots are in nucleus. B. GFP-mMBD4 and FLAG-PRMT5 were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots that are associated with heterochromatic chromatins. About half of the double transfected cells show changes at MBD4 highly expressing chromatins, Surrounding the PRMT5 spots in nucleus (See C & also see supplementary Figure. F1&F2).

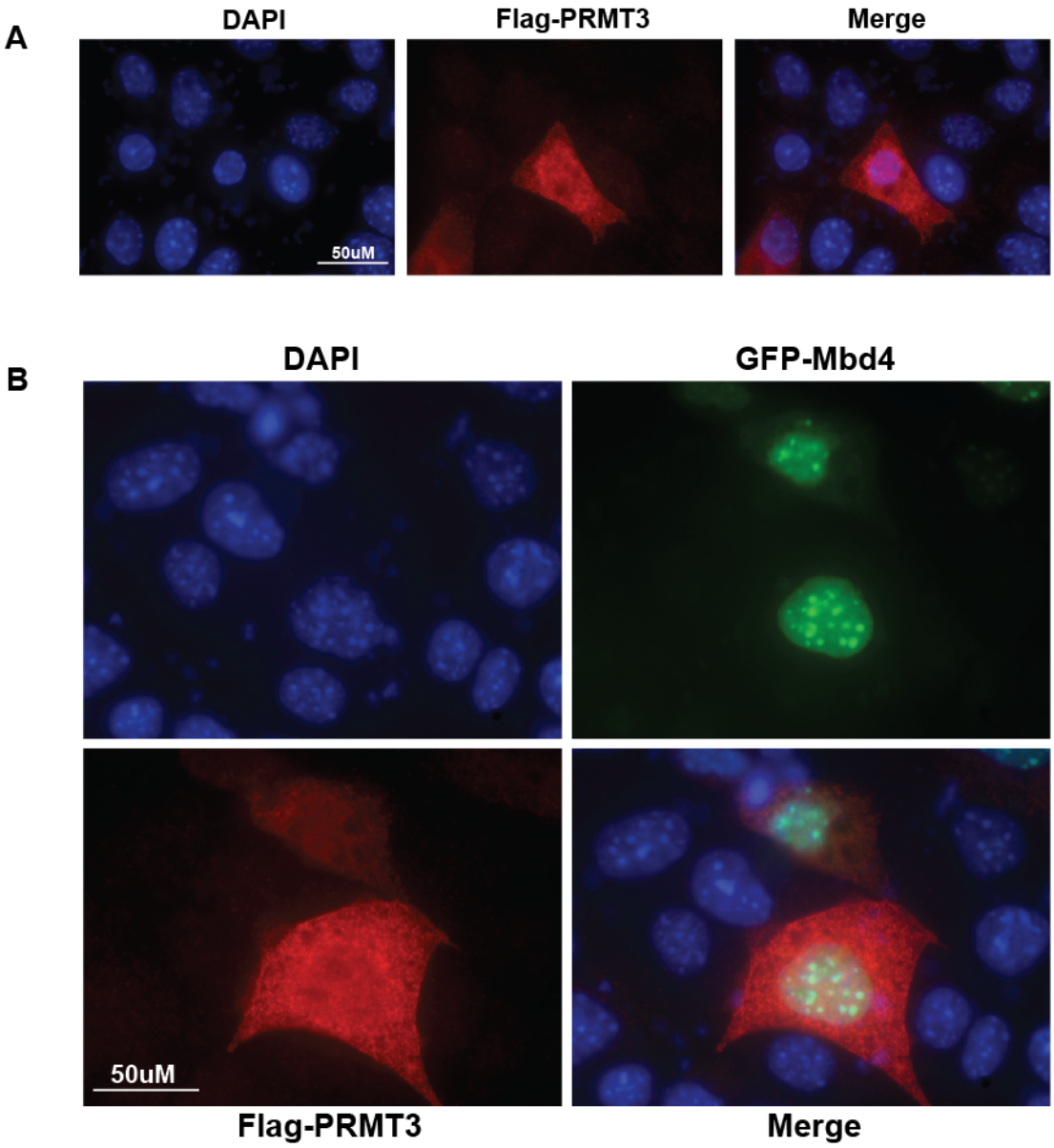


Figure. 4.9. PRMT3 shows ubiquitous distribution in the whole cell body and colocalises with MBD4 and bright DAPI spots. A. Distribution of PRMT3 was compared with DAPI counterstaining. Vectors containing FLAG-PRMT3 were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed PRMT3 exhibits diffused general distribution in whole cell. The bright DAPI spots were colocalised with such diffused distribution of PRMT3. B. GFP-mMBD4 and FLAG-PRMT3 were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots, whereas PRMT3 shows diffused general distribution the same as controls, but No bright spots of PRMT3 were observed. All the cells single- or double transfected show this pattern (See supplementary Figure. 4.9_4.10 G1&G2)

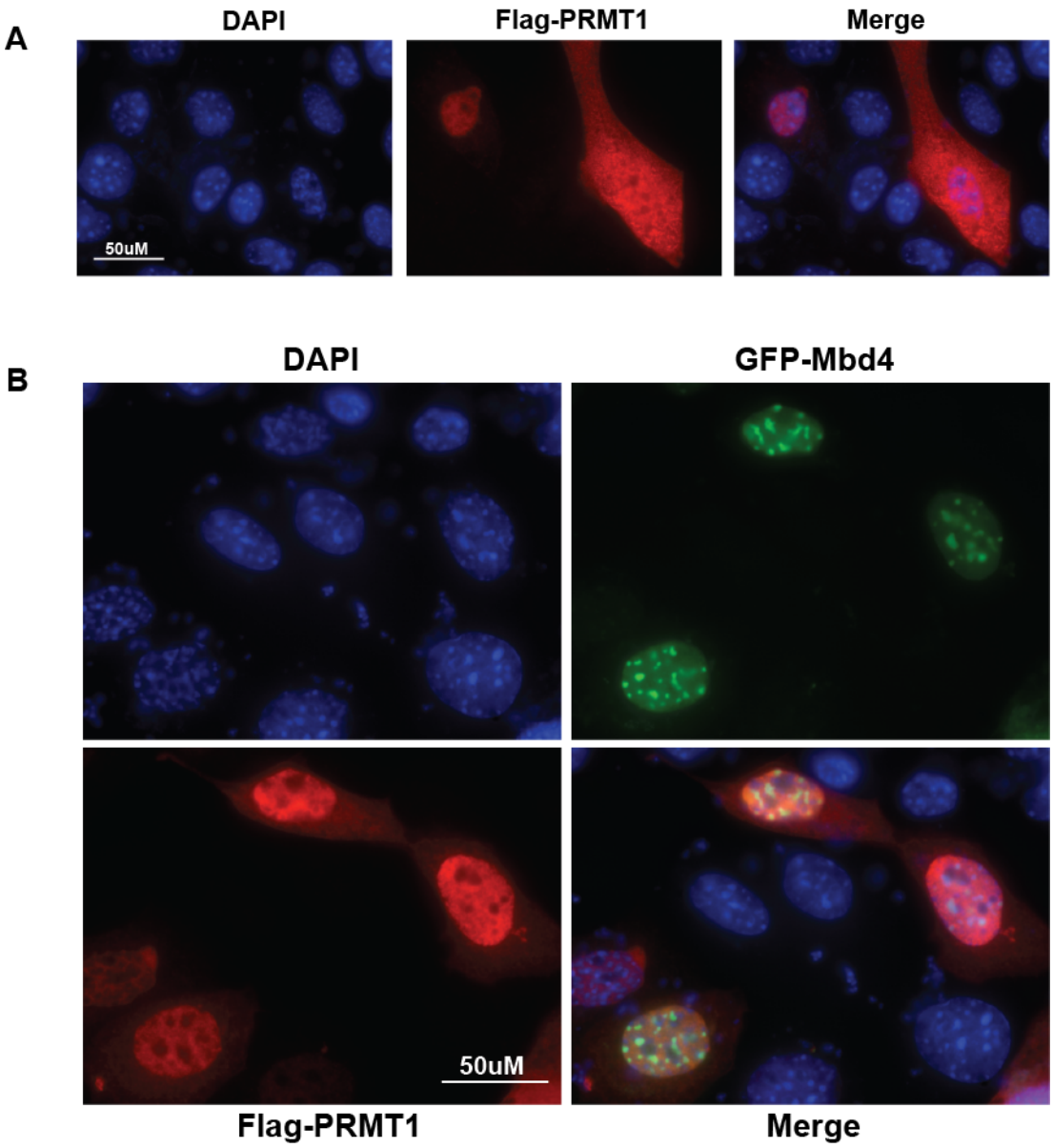


Figure. 4.10. PRMT1 shows a diffused distribution in both cytoplasm and nucleus, partially colocalises with MBD4 and bright DAPI spots. A. Distribution of PRMT1 was compared with DAPI counterstaining. Vectors containing FLAG-PRMT1 were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed PRMT1 exhibits diffused general distribution in the whole cell. The bright DAPI spots were partially colocalised with such diffused distribution of PRMT1. B. GFP-mMBD4 and FLAG-PRMT1 were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots, and PRMT1 shows diffused distribution mainly in the nucleus, and partially colocalises with DAPI and MBD4 bright spots. All the cells single- or double transfected show this pattern (See supplementary Figure. 4.9_4.10 H1&H2).

Similar to their endogenous expression (Lee and Skalnik, 2002), overexpressed FLAG-CFP1 and FLAG-SETD1a proteins were localized to the nucleus in CMT93 cells and exhibits a punctate or sparkle distribution (Figure. 4.11-4.12 A). These speckles are concentrated in areas of DAPI light staining and are clearly excluded from the DAPI and MBD4 bright spots, consistent with their localizations to euchromatic regions of the nucleus (Lee and Skalnik, 2002). Some sparkles or dense staining of FLAG-CFP1 can be observed in cytoplasm, and the same was observed in several SETD1a transfected cells. Notably, their distribution only partially overlapped with co-transfected GFP-MBD4. The MBD4 distribution was largely not affected when cotransfected with CFP1 or SETD1a in CMT93 cells (Figure. 4.11-4.12 B).

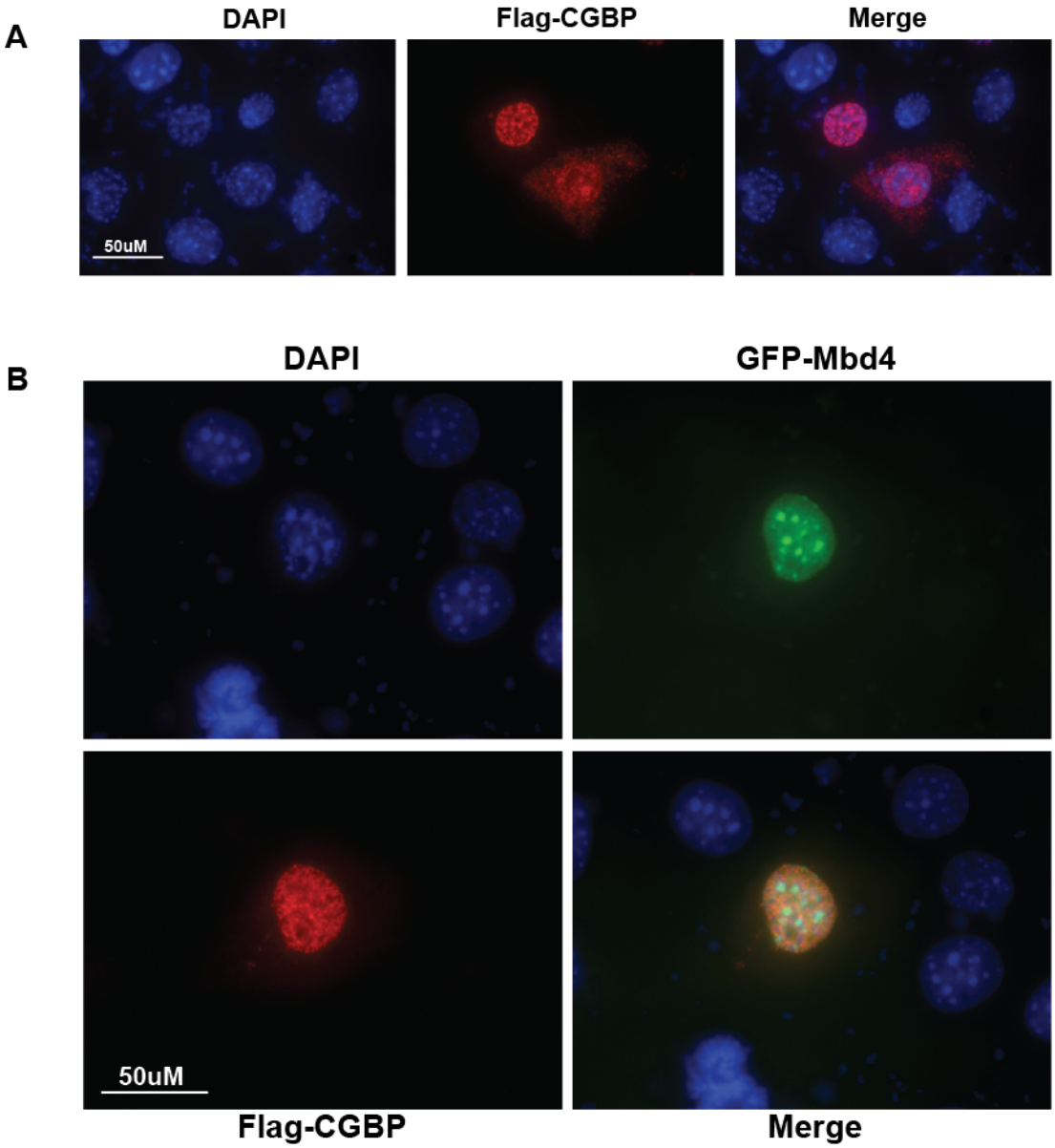


Figure. 4.11. CGBP (CFP1) localises to nucleus speckles in MBD4 and DAPI light regions. A. Nucleus distribution of sparkle pattern of FLAG-CGBP fusion proteins (like endogenous CGBP protein) was detected in CMT93 cells using anti-FLAG antibody and Texas Red-conjugated secondary antibody. Fixed cells were counterstained with DAPI and observed by fluorescent microscopy. Vectors expressing FLAG-tagged CGBP were transiently transfected into CMT93 cells. B. Nucleus distribution of speckles containing GFP-CGBP was compared with GFP-mMBD4 and DAPI staining in CMT93 cells. Overexpressed MBD4 colocalises with DAPI bright spots, and CGBP shows sparking distribution mainly in nucleuss, and partially colocalises with DAPI and MBD4 bright spots. All the cells single- or double transfected show this pattern (See supplementary Figure. 4.11_4.12 I1&I2).

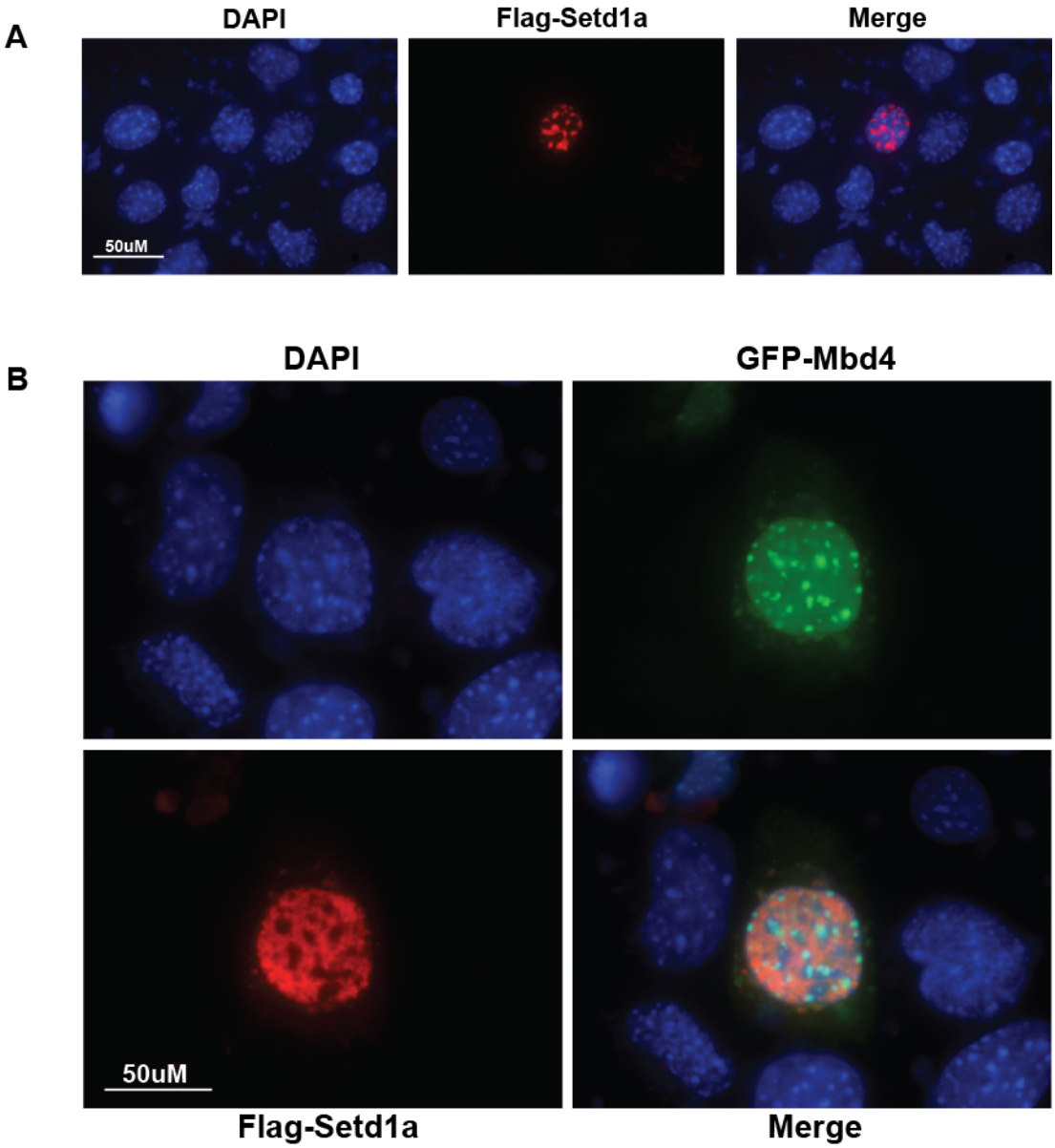


Figure. 4.12. SETD1a shows diffused distribution in nucleus, excluded from MBD4 and bright DAPI spots. A. Distribution of SETD1a was compared with DAPI counterstaining. Vectors containing FLAG-SETD1a were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed SETD1a exhibits diffused general distribution in nucleus. The bright DAPI spots were excluded from such diffused distribution of SETD1a. B. GFP-mMBD4 and FLAG-SETD1a were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots, and SETD1a shows diffused distribution mainly in nucleus, and excluded from DAPI and MBD4 bright spots (More pictures see supplementary Figure. 4.11_4.12 J1&J2).

DEK and SRSF2/SC35

Overexpressed FLAG-DEK and GFP-SC35 exhibit overall diffused distribution in nucleus, but with nucleoli excluded from their distribution. Most of the DAPI bright spots were excluded from the FLAG-DEK and GFP-SC35 distribution (Figure. 4.13-4.14 A). when co-transfected with MBD4, a small fraction of MBD4-containing spots co-localized with the RNA splicing factor GFP-SC35 and oncogene FLAG-DEK. However, It is noteworthy that the overall distribution of MBD4 is clearly distinct from that of GFP-SC35 and FLAG-DEK (Figure. 4.13-4.14 B).

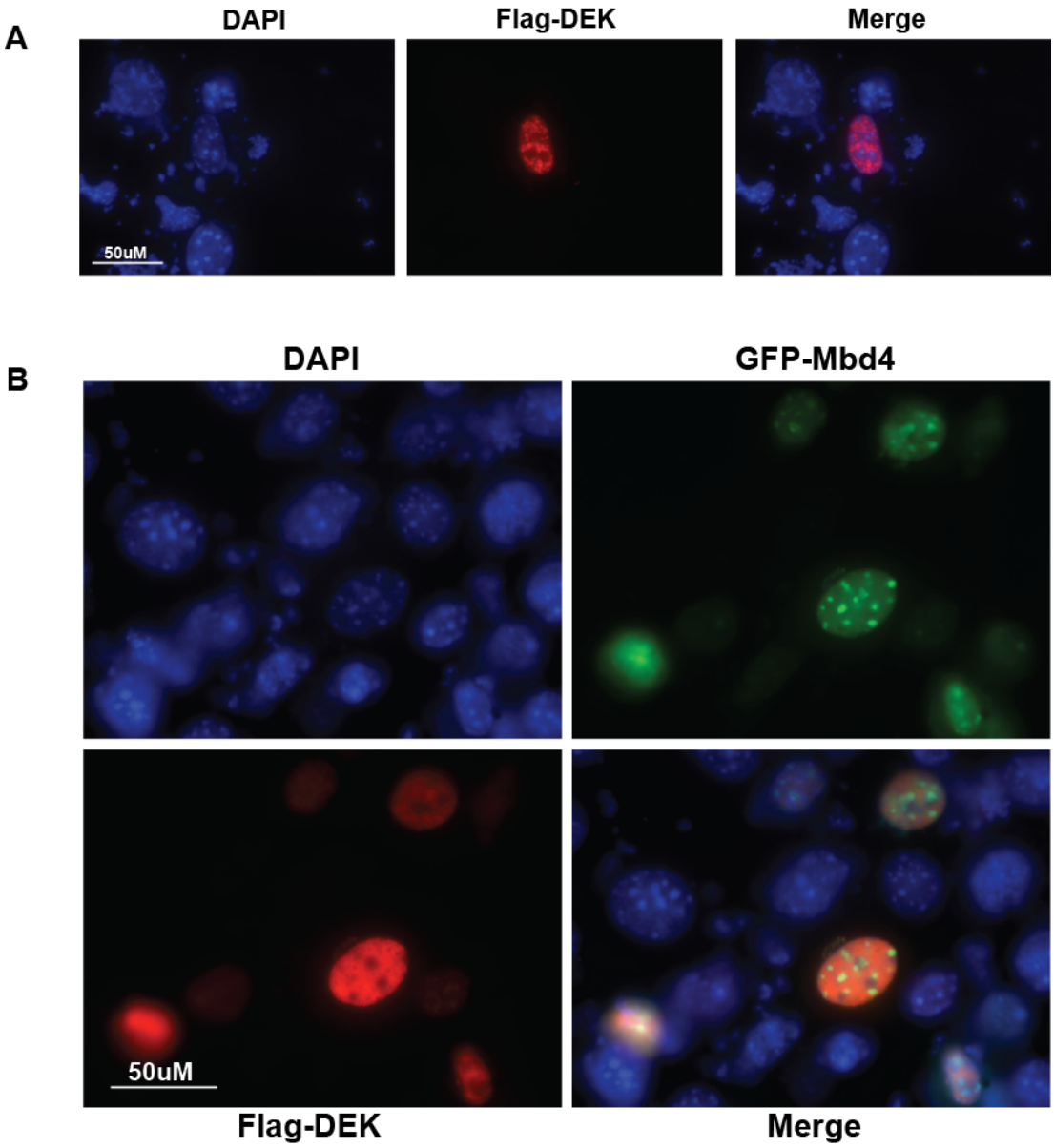


Figure. 4.13. DEK shows a diffused distribution in nucleus, partially overlaps with MBD4 and bright DAPI spots. A. Distribution of DEK was compared with DAPI counterstaining. Vectors containing FLAG-DEK were transiently transfected into CMT93 cells, and were detected with anti-FLAG antibody and Texas Red-conjugated secondary antibody. The overexpressed DEK exhibits diffused general distribution in nucleus. The bright DAPI spots were partially overlap with such diffused distribution of DEK. B. GFP-mMBD4 and FLAG-DEK were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots. DEK shows diffused distribution in nucleus, and partially overlap with DAPI and MBD4 bright spots. All the cells single- or double transfected show this pattern (See supplementary Figure. 4.13_4.14 K1&K2).

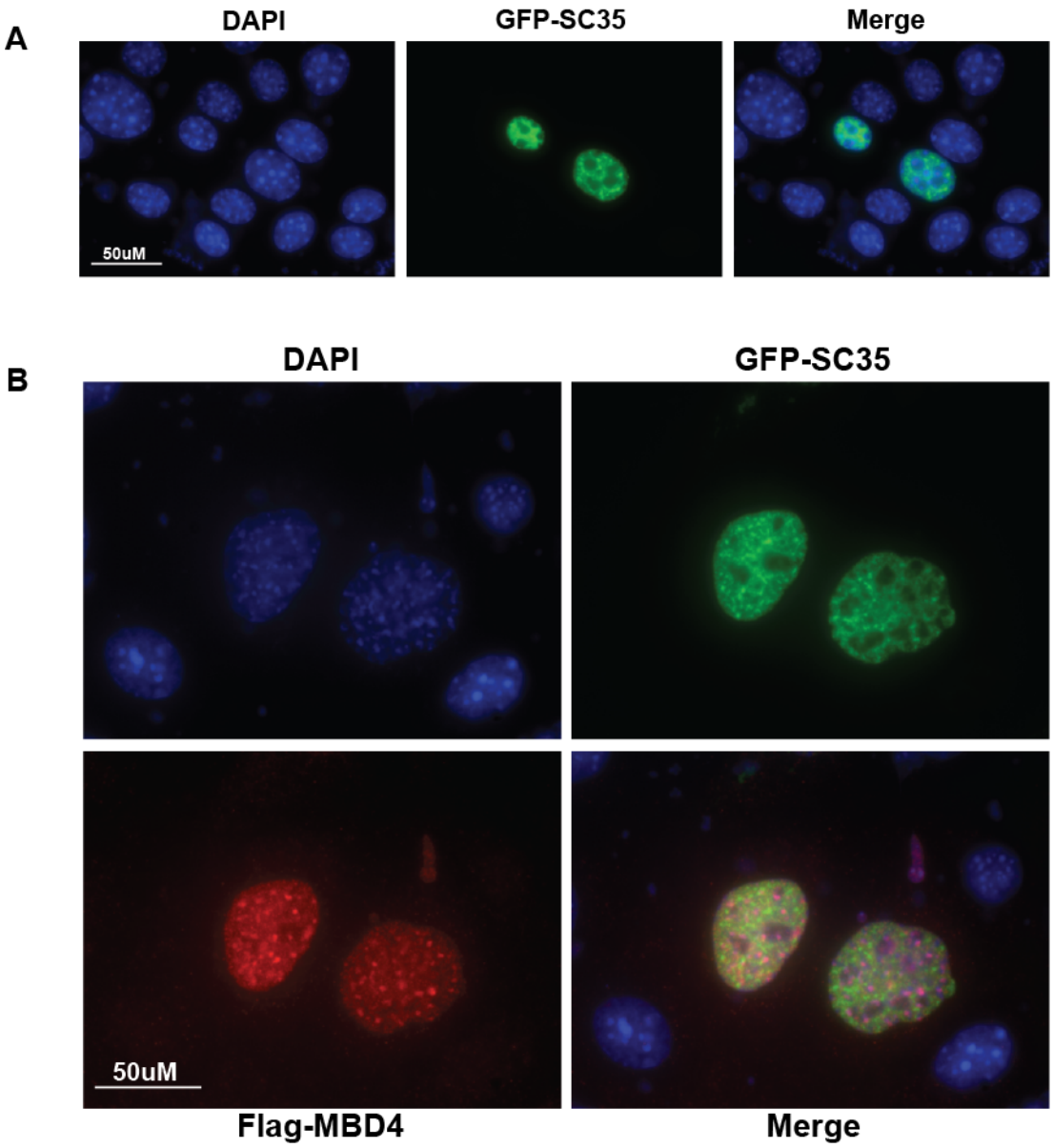
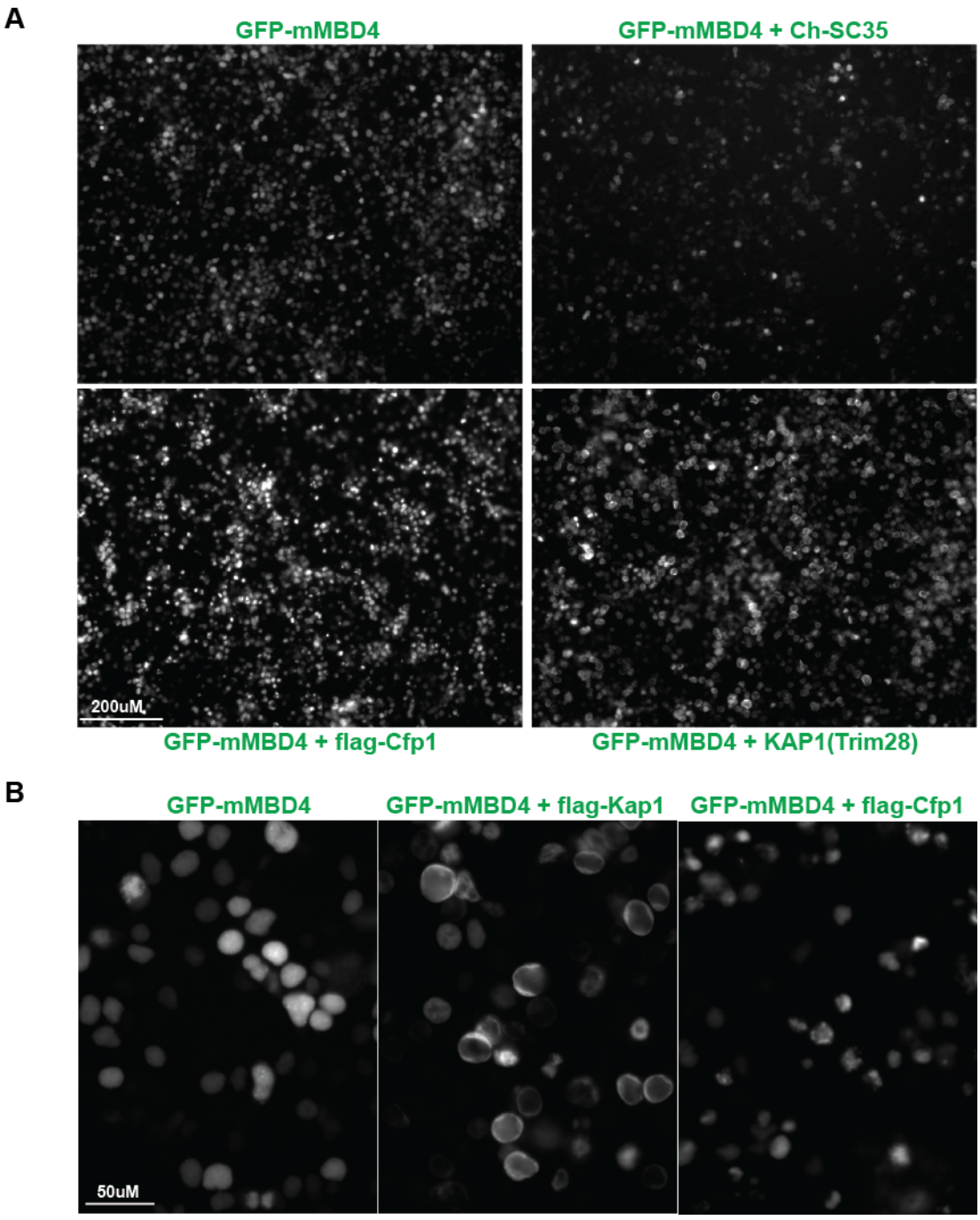


Figure. 4.14. SC35 shows diffused distribution in nucleus, partially overlaps with MBD4 and bright DAPI spots. A. Distribution of SC35 was compared with DAPI counterstaining. Vectors containing GFP-SC35 were transiently transfected into CMT93 cells. The overexpressed SC35 exhibits diffused general distribution in nucleus. The bright DAPI spots were partially overlapping with such diffused distribution of SC35. B. FLAG-hMBD4 and GFP-SC35 were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots, and SC35 shows diffused distribution with sparkling in nucleus, and partially overlap with DAPI and MBD4 bright spots. All the cells single- or double transfected show this pattern (See supplementary Figure. 4.13_4.14 L1&L2).

4.2.3. Some MBD4-associated Proteins affect the nucleus localization and expression intensity of MBD4 when over-expressed

By IP and IF, I have validated several robust MBD4-interacting proteins such as UHRF1, USP7, and ACF1. However, there are some candidates in my validation list can be immunoprecipitated but their distribution patterns were largely distinct from that of MBD4. I then addressed if the expression level or pattern of MBD4 could be affected by co-transfection of these candidates. I co-transfected GFP tagged MBD4 and the chosen candidates into HEK293T cells. Surprisingly, although the nucleus distribution of KAP1, CFP1 and SC35 were entirely distinct from that of MBD4 in mouse CMT293T cells, they did indeed affect both expression intensity and distribution pattern of GFP-MBD4 in 293T cells. Overexpressed KAP1 and CFP1 can largely increase the expression intensity of cotransfected GFP-MBD4, whereas overexpressed SC35 and USP7 decreases the intensity of GFP-MBD4 (Figure 4.15). The distribution patterns of MBD4 were completely changed by the cotransfections of FLAG tagged KAP1 and CFP1, concentrating at the nucleus periphery in cotransfection with KAP1, while becoming very condensed in the case of CFP1. It is noteworthy that my reciprocal co-IPs were performed in 293T cells. In CMT93 cells, I could not observe such changes. However, at least in 293T cells, it is very clear that these MBD4 associated proteins including KAP1, CFP1, USP7, and SC35 affected the nuclear localisation and expression intensity of GFP-MBD4 (Figure. 4.15).


A
GFP-mMBD4
GFP-mMBD4 + Ch-SC35
200uM
GFP-mMBD4 + flag-Cfp1
GFP-mMBD4 + KAP1(Trim28)
B
GFP-mMBD4
GFP-mMBD4 + flag-Kap1
GFP-mMBD4 + flag-Cfp1
50uM

C

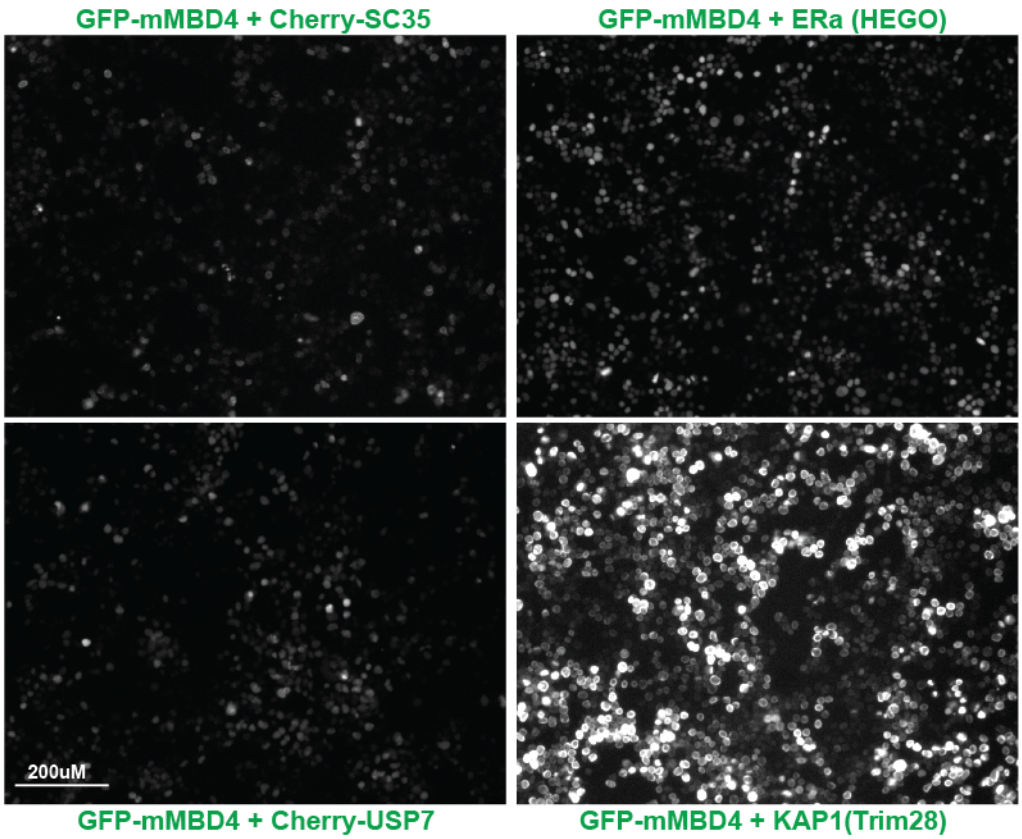

GFP-mMBD4 + Cherry-SC35
GFP-mMBD4 + ERa (HEGO)
200uM
GFP-mMBD4 + Cherry-USP7
GFP-mMBD4 + KAP1(Trim28)

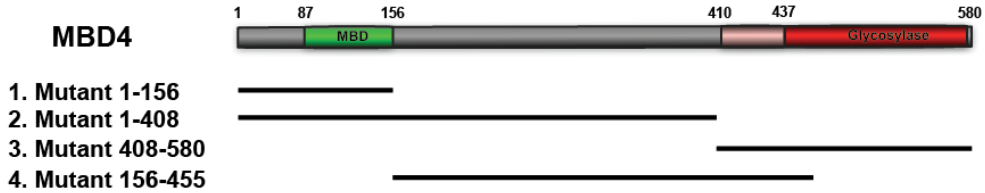
Figure. 4.15. MBD4 associated protein candidates can influence the expression intensity and pattern of MBD4. The GFP-tagged MBD4 was transiently transfected into 293T cells singly (A. upper left) or co-transfected with respective interacting candidates. (A. upper right) SC35 decreases the expression intensity of MBD4 at ~26%, in contrast to CFP1 (A lower left, increase ~28%) and KAP1 (A. lower right, increase ~15%) that increases or stabilises MBD4 expression. (B) Single transfection of MBD4 shows a ubiquitous expression within nucleuss in 293T cells, with little bright spots inside (Left). In contrast, co-transfection of KAP1 results in the accumulation of MBD4 expression at nucleus peripheral (middle), and MBD4 shows a condensed distribution of MBD4 within nucleus when co-transfected with CFP1 (right). (C) In a certain concentration, the intensity of MBD4 can be influence dramatically by co-transfection of its associating protein candidates. In this batch of transfections, SC35 decreases the expression of MBD4 at similar level (upper left), in contrast to KAP1 that increases and stabilises MBD4 expression dramatically at this time at ~ 282% (lower right). Co-transfections of ERα (upper right) was similar to the single transfection of MBD4 (A upper left). USP7 (lower left) decreased MBD4 expression at ~33%. All the pictures were taken at full confluence and where possible at the regions that no overlapping cells were observed. Mean Intensities were calculated by ImageJ, in which the total intensity of the picture was divided by the total area of the picture. All the mean intensities were normalised to the picture of double transfection of MBD4 and SC35. As I chose full confluence regions in all the transfections to take picture, I assume the total number of cells in each picture were roughly similar. Thus the ratios of intensity changes can be considered as a rough measurement to give a brief idea how much changes in each pictures compared with the GFP-MBD4 single transfection control (Millar, 2002).

4.2.4. Identification of the interacting domains of MBD4 responsible for candidate binding by pull down assays.

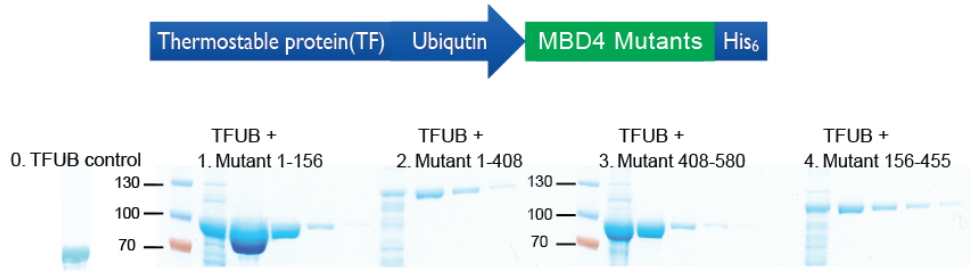
In order to define the interacting domains within MBD4 responsible for its associations with the identified MBD4 interacting proteins, I made different mutants of MBD4 that were extracted and purified in vitro from *E.coli*. Subsequently I characterised the interacting domains of MBD4 responsible for the associations with the identified interacting proteins by reciprocal pull down assays. Four truncated versions of MBD4 (Figure. 4.16A) were generated and expressed in *E.coli* as HIS-tagged fusion proteins (Figure. 16B). The HIS-tagged fusion mutants have a thermostable protein called trigger factor as well as ubiquitin at its upstream (Figure. 4.16B), which make the fusion proteins very soluble and stable in the following extraction and purification processes (Thapa et al., 2008). The four truncated mutants of MBD4 were designed to represent the MBD, interveining, and glycosylase domains of MBD4, and part of these mutants overlap with each other to minimise the interacting regions that were specifically responsible for the associations. The purified His-MBD4 mutants were then reciprocally co-immunoprecipitated with the identified MBD4 interacting proteins with FLAG- or GFP- tags that were overexpressed in HEK 293T cells. I

used MLH1 as the positive control, as it robustly associates with MBD4, and its interacting domains had been fine mapped (Millar, 2002). The interacting domains that were identified by my reciprocal pull downs were consistent with the previous findings, showing the robustness of my mutants pull down approach (amino acids (AAs) 410-455 from my assays in Figure. 4.16 D & AAs 415-420 from Catherine Millar Thesis (Millar, 2002), Adrain Bird Lab). Finally I identified the interacting domains within MBD4 responsible for its associations with UHRF1, PRMT5, and KAP1. Notably I characterised the interveining domain of MBD4 were largely responsible for the MBD4 associations of UHRF1, PRMT5, and KAP1. This region within MBD4 was previously reported as a function desert, my characterisation here elucidates that this interveining region may possess a significant role for protein interactions, which might be independent from its glycosylase repair and DNA binding function.

A



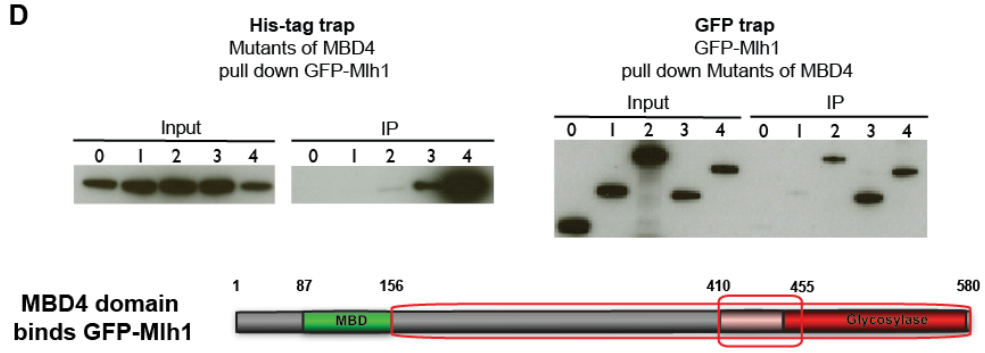
B



C

His-tag trap	Reciprocol flag or GFP trap
6HisTFUB-Mutants purified from E-coli **Pre-incubate to beads overnight** **+** **Extensive 293T cell lysate** (overexpressed cadidate protein) for 1 hour cold room rotating	6HisTFUB-Mutants purified from E-coli **pre-incubate to extensive 293T cell lysate** (overexpressed cadidate protein) overnight **+** **Beads** for 1 hour cold room rotating
Extensive washes by **Wash** buffer or **RIPA** buffer Beads immediately boil in loading buffer	

D



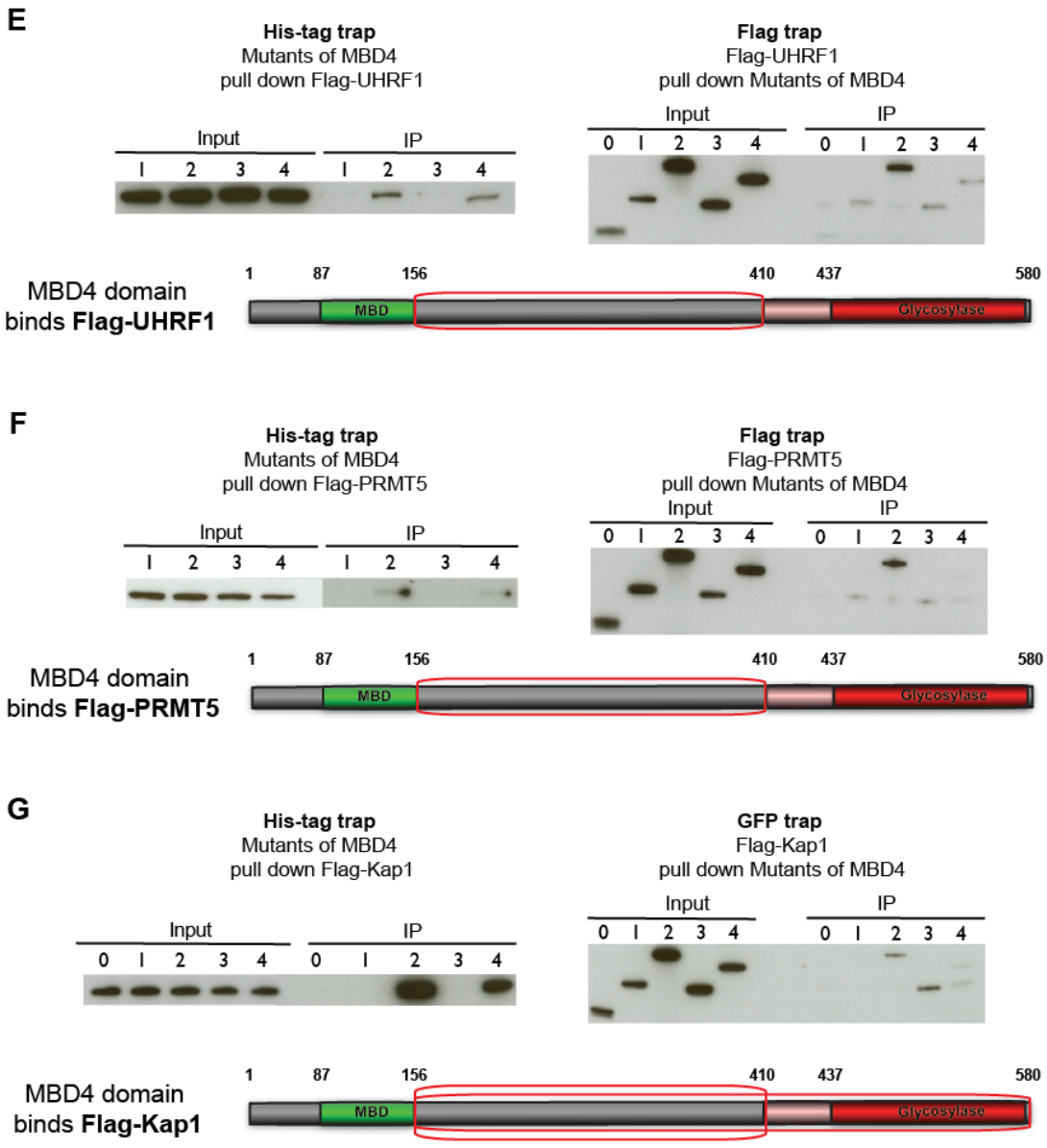


Figure. 4.16. Identification of interacting domains of MBD4 responsible for the associations with its confirmed interacting candidates UHRF1, PRMT5, and KAP1. (A) The scheme of four mutants representing different domain of MBD4. (B) the schematic representation of the fusion protein mutants of MBD4, and coomassie results of the purified MBD4 mutants fused with trigger factor and ubiquitin fusions (6His-TF-UB). (C) Flow chart of the reciprocal co-IP procedure. (D) Reciprocal co-IPs for identification of interacting regions of MBD4 for its association with MLH1. the region labelled by red square is the interacting domain identified, which correlates with the previous reported GST pull down findings (Catherine Millar Thesis, Adrain Bird Lab), and served as positive control. (E) (F) (G) the interacting domains of MBD4 responsible for the associations of UHRF1, PRMT5, and KAP1 are at its interveining region.

4.3. Discussion

In this chapter, I selected a number of MBD4 interacting protein candidates to verify their association with MBD4 by reciprocal co-IP and IF colocolisation. The majority of positive results not only confirmed the robustness of my Y2H and AP/MS screens, but also provided me with several robust MBD4 partners for future investigations. These include USP7, UHRF1, ACF1, PRMT5, and KAP1. Here I discuss the possible implications of these MBD4 protein interactions, which include the regulation of DNA methyltransferases activity through ubiquitination and deubiquitination of UHRF1 and USP7; induction of transcriptional repression by chromatic remodelling of ACF1 and histone modification of PRMT5 and by recruiting transcriptional co-repressor KAP1. All of these may facilitate the functions of MBD4 in transcriptional repression, DNA repair, apoptosis, and in the possible active DNA demethylation pathways.

Here I need make it clear that the colocalisation experiments on over-expressed tagged proteins may "over-represent" protein interactions because the level of the proteins are not physiological. Thus the role of MBD4 in regulating DNA methylation (through UHRF1), remodelling (Through ACF1), repression (Through KAP1) etc. are really speculations/hypothesises which I would have tested had there been sufficient time, for example, the assays of methylation in *Mbd4*$^{-/-}$ cells using HPLC followed by next generation sequencing etc.

4.3.1. Regulation of stability of DNA methylation: a possible role of MBD4 through its association with USP7 & UHRF1?

USP7 (Ubiquitin specific peptidase 7, herpes virus-associated; also known as HAUSP) is a deubiquitinating enzyme, a member of the ubiquitin specific peptidase class of deubiquitinating enzymes (DUBs) (Qin et al., 2011). Depletion of *Usp7* in mice results in arrest of embryonic development shortly after implantation (E6.5–7.5) (Kon et al., 2010), and its associating developmental factors and substrates including EBNA1 (Holowaty, 2003), DAXX (Tang et al., 2006), FOXO4 (van der Horst et al., 2006), PTEN (Song et al., 2008), and loss of their functions may contribute to this early embryonic lethal phenotype. USP7 takes part in the cellular stress responses by regulating the stability of p53 and MDM2 through its deubiquitination activity (Li et al., 2002, Cummins and Vogelstein, 2004, Cummins et al., 2004, Meulmeester et al., 2005). It has been proposed that USP7 may have dual roles; In addition to a direct deubiquitylation of histones in chromatin remodelling,

USP7 may regulate the cellular levels of E3 ubiquitin ligases indirectly involved in histone ubiquitylation (Khoronenkova et al., 2011). In particular, the deubiquitination of histone H2B can be enhanced by the direct interaction of GMP-synthetase and USP7, which act as a transcriptional corepressor (van der Horst et al., 2006, van der Knaap et al., 2005, Sarkari et al., 2009). Furthermore, it has been shown that USP7 can modulate MDM2 cellular levels and significantly reduce chromatin DNA accessibility and the rate of repair of oxidative lesions, while no change of the levels or activity of base excision repair enzymes was observed in the USP7 transient siRNA knockdown. This suggests that USP7 may modulate chromatin remodeling that recruits a repressive chromatin environment suitable for efficient base excision repair (BER) of oxidative lesions (Khoronenkova et al., 2011).

UHRF1 (Ubiquitin-like with PHD and ring finger domains 1; also known as ICBP90 in human; NP95 in mouse) plays a central role in the maintenance of genomic DNA methylation. UHRF1 interacts with and recruits DNMT1 to hemi-methylated DNA in daughter DNA strands to faciltate their methylation. In addition, the SRA domain of UHRF1 not only recognises and binds preferentially to hemi-methylated DNA, but also has a binding specificity to newly discovered 5-hydroxymethylcytosine (5hmC) probes with similar affinity to 5-methylcytosine probes by a putative flip mechanism (Frauer et al., 2011, Rajakumara et al., 2011, Sharif et al., 2007). DNMT1 has such flip mechanism as well (Song et al., 2012). UHRF1 also binds to di- and tri-methylated histone H3K9, as well as ubiquitination of histone H3, and associates with heterochromatin formation, suggesting an important role of UHRF1 in linking histone modifications, DNA methylation, and chromatin structure (Unoki et al., 2009). The UHRF1 complex includes a cohort of crucial developmental factors including HDAC1, Tip60, G9a, and maintenance and de novo DNMTs, thus it is not surprising that inactivation of mouse NP95 (mouse version of Uhrf1) was lethal to midgestation embryos (Muto, 2002). UHRF1 is upregulated in variety of cancers, which is probably associated with the the tumor suppressor p53, as it was shown that UHRF1 may be downregulated by p53, which is lacking in 50% of cancers, through up-regulation of p21/WAF1 and subsequent deactivation of E2F1 (Unoki et al., 2009a, Unoki et al., 2009b).

In this chapter, I have shown that MBD4 interacts with USP7 and UHRF1 directly by reciprocal co-IP and IF. Interestingly co-transfection of MBD4 dramatically changed the nucleus localisation of USP7, such that MBD4 recruits USP7 to heterochromatic sites, which coincide with the similar observation of our previous finding of the recruitment of MLH1 to micro-irradiated heterochromatic sites led by MBD4 (Ruzov et al., 2009). In addition, co-introduction of MBD4 and UHRF1 not only co-localises tightly together in the nucleus, but

also causes chromatic changes, and induces the drastic reduction of the DAPI bright spots, where are generally associated with heterochromatic chromatins. It has been shown that a large macromolecular protein complex containing DNMT1, USP7, TIP60, UHRF1, HDAC1, and PCNA accumulates and binds at the DNA replication fork and the components of this complex co-ordinately regulate the functional availability of DNMT1 (Du et al., 2010, Qin et al., 2011, Felle et al., 2011). My findings of direct interactions between MBD4 and USP7 and UHRF1 have suggested a previously unknown pathway that MBD4 might co-ordinately modulate the ubiquitination of DNMT1, and consequently regulates the stability and enzymatic activity of DNMT1 (Figure. 4.17). Furthermore, we have previously shown that MBD4 and DNMT1 could localise to sites of DNA damage, together with PCNA and MMR protein MLH1 (Ruzov et al., 2009). Additionally, recent work suggests that DNMT1 participates in the MMR pathway and can be recruited to sites of DNA damage by PCNA (Loughery et al., 2011, Guo et al., 2004, Mortusewicz et al., 2005). These data suggest that the interaction of these partners on chromatin is influenced by DNA damage and that, while present at these lesions, it is possible that MBD4 and DNMT1 can participate in the decision to either repair the DNA damage, maintain the DNA methylation profile, or activate a cell death program (Ruzov et al., 2009). In general, at the initiation of DNA replication, MBD4 might be involved in stabilizing the maintenance methyltransferases DNMT1, by firstly regulating the ubiquitination activity of UHRF1 to ubiquitinate DNMT1, thereby leading to proteasomal degradation, and subsequently recruits USP7 and HDAC1 to the established complex containing DNMT1 and UHRF1, and localise the macromolecular complex to the replication forks through the interaction with PCNA (Figure. 4.17. A). The components of the accumulating complex can be maintained at a proper level by the concerted actions of the ubiquitination and de-ubiquitination activities of UHRF1 and USP7, which may be regulated by MBD4, so that the methylation maintenance can be completed by the appropriate activity of DNMT1. In addition, during and after the cell cycle where DNA replication errors occur including the DNA damages and DNA methylation mismatches, MBD4 itself as a glycosylase, together with the MMR proteins such as MLH1, may recognise, excise and repair the DNA damages and maintain the DNA methylation profile, or induce apoptosis if the errors are fatal (Figure. 4.17. B). Finally, at the end of DNA replication when the entire hemi-methylated DNA is re-methylated, the excess amounts of DNMT1 can be sensed and ubiquitination-mediated proteasomal degradation of DNMT1 is triggered (Du et al., 2010). In this model, MBD4 may directly interact with the key components of the complex, and might regulate the sequential posttranslational events at the DNA replication fork, recruit a repressive chromatic environment and act as a co-repressor, repair and maintain the

methylation profile or trigger the apoptosis pathway, contributing to the regulation of DNMT1 stability during the cell cycle. Lastly, it is possible that the recently discovered 5-hydroxymethylcytosine could be recognised and contribute to this process, as the SRA domain of UHRF1 in this complex may bind to the hydroxy-marks (Frauer et al., 2011), and MBD4 is a well-known glycosylase in the base excision repair pathway, which also interacts with MMR proteins such as MLH1. It is noteworthy that MBD4 by itself cannot bind to hydroxymethylated DNA at least in my in vitro EMSAs in chapter 6. However, it does not exclude the possibility that a glycosylase like MBD4 in a large protein complex containing some hydroxymethylation-binding proteins might be involved in this novel pathway, and/or MBD4 in such complex may recognise and excise other hydroxy-products of the possible DNA demethylation pathways (Hashimoto et al., 2012). My finding of their direct interaction suggests the complex containing MBD4, UHRF1 and DNMT1 may be an important candidate responsible for the putative demethylation process mediated by either the canonical model of thymine conversion by deamination of 5mC followed by glycosylase excision and BER pathway repair, or the possible hydroxy-methylation route.

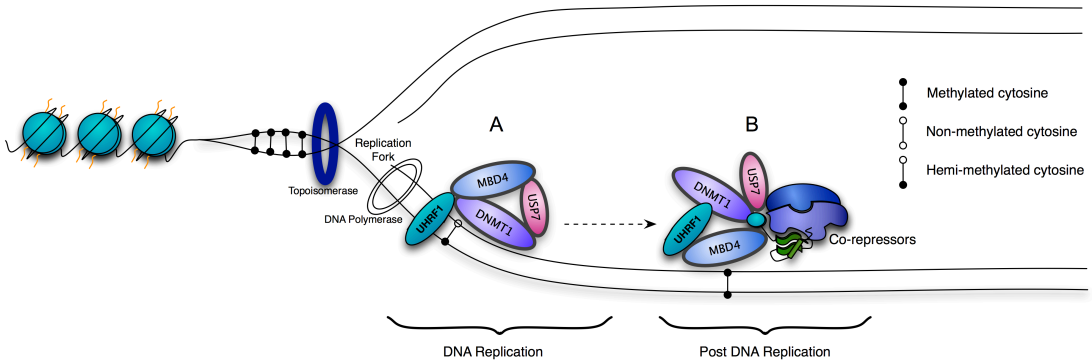


Figure. 4.17. A working model for the role of MBD4 in regulation of stability of DNA methylation. (A.) MBD4 may have a role in regulation of methylation machinery during the DNA replication, through its direct interaction with DNMT1 (previously shown), UHRF1 (novel) and USP7 (novel). (B.) In addition, MBD4 may be responsible for the maintenance of fidelity of methylation marks at specific loci after DNA replication, by its glycosylase activity and its ability of recruitment of the components of methylation machinery, together with recruiting the co-repressors complexes.

4.3.2. Role of MBD4 in nuclear remodelling through ACF1?

ACF1 (ATP-utilizing chromatin assembly and remodelling factor 1, also known as BAZ1A) is the one of protein components in the complexes of ACF, WCRF and CHRAC, all of which are conserved in *Drosophila melanogaster*, *Xenopus laevis* and human cells (Langst

and Becker, 2001). ACF1 is non-catalytic, however, it can stimulate the catalytic activity of the other proteins in the ACF complex, the ATPase motor protein SNF2H (sucrose nonfermenting-2 homolog, also called SMARCA5) (Ewing et al., 2007), and may guide the enzymatic components to its target. The catalytic function of ACF complex is required in both the histone deposition and the relaxation of chromatin structure (Ewing et al., 2007, Ito et al., 1999, Collins et al., 2002). Nucleosome remodelling, which renders nucleosomal arrays dynamically driven by ATPase to disrupt histone-DNA contacts, increases the accessibility of nucleosomal DNA and induces histone octamer translocation on DNA (Narlikar et al., 2002, Eberharter et al., 2004, Becker and Horz, 2002, Lusser and Kadonaga, 2003). It has been shown that ACF1 associates with the nucleosome remodelling ATPase ISWI, improves the efficiency of ISWI-induced nucleosome sliding in replication and transcription (Eberharter et al., 2004). Together with the other component SNF2H of the ACF complex, ACF1 may act as a key regulator in the nucleosome assembly and contribute to transcriptional repression of heterochromatic genes by maintaining the chromatin architecture (Ewing et al., 2007).

A growing body of evidence suggests that remodelling factors are recruited to specific nucleosomes by site-specific DNA-binding proteins (Narlikar et al., 2002, Peterson and Logie, 2000, Li et al., 2004). One example would be the vertebrate Mi-2/NuRD (nucleosome remodeling and deacetylase) complex, which contains histone deacetylases (HDAC1/2) and methyl-CpG binding domain (MBD) proteins, in particular, MBD3 (Clapier and Cairns, 2009). In this chapter, I identified that MBD4 interacts and colocalise with ACF1 at heterochromatic sites, and they exhibit concentrated enlarged heterochromatic bright dots in cell nucleus. This suggests another possible link between methyl-CpG binding proteins and chromatin remodelling factors. Interestingly, as mentioned above, MBD3 is also a candidate from my AP/MS screen, which potentially interacts with MBD4 directly. Additionally, ACF1 and its complexes have been shown to play important roles in DSBs repair (Lan et al., 2010). Notably, one of the DSBs repair protein MRE11 was also screened out in my AP/MS screen. These data suggest the interaction of MBD4 and ACF1 may be required for establishment of repressive environment by chromosome remodelling, for the effective repair including BER pathway and possibly DSBs repair pathway. These repairs might be involved in one of the possible demethylation pathways. One of the clues is that a novel regulatory role for ACF1 recently discovered in transcriptional repression of Vitamin D3 receptor-regulated genes (Ewing et al., 2007). This study shows that hACF1 affects the hormone responsiveness and the transcriptional repression of specific class II NR-regulated genes, and demonstrates that hACF1 is critical for the stable occupancy of the N-CoR

corepression complex at the target gene promoters (Ewing et al., 2007). co-incidentally MBD4 was reported to be involved in a possible DNA demethylation in hormone-induced transcriptional derepression at a specific Vitamin D responsive promoter, in a negative vitamin D response element (nVDRE) binding protein complex containing two DNA methyltransferases, DNMT1 and DNMT3B (Kim et al., 2011) (Finally this paper was withdrawed by some of the contents were interesting). The coincidence of the roles of ACF1 and MBD4 at Vitamin D regulated genes suggest that ACF1 might remodel the nucleosomes at heterochromatin sites and make them accessible to let repair proteins such as MBD4 to target the specific methylated gene promoters, and subsequently switch to transcriptional derepression through a possible demethylation mechanism, in response to the hormone treatment. It is noteworthy that the other interacting partners of MBD4 including UHRF1 and the possible partner MBD3 were implicated in the newly discovered 5-hydroxymethylation recognition, albeit the evidence of their specificity is not strong enough (Yildirim et al., 2011). In theory, the 5-methylcytosine to 5-hydroxymethylcytosine transitions at gene promoter regions may also induce the putative demethylation, thus it is very interesting to further investigate these MBD4 interactions in future functional studies.

KAP1

KAP1 ((Kruppel-associated box)-associated protein 1, also known as TRIM28; TIF1β) is a scaffold protein that can assemble epigenetic machinery (Iyengar and Farnham, 2011). Kap1 has been shown to be involved in multiple cellular processes including gene silencing, cell growth and differentiation, pluripotency, neoplastic transformation, apoptosis, DNA repair, and maintenance of genomic integrity (Iyengar and Farnham, 2011). KAP1 may impact epigenetic patterns and chromatin compaction, and subsequently regulate the dynamic organization of chromatin structure, through its genomic targeting ability, protein-protein interactions, and specific post-translational modifications of itself and its associated histones (Iyengar and Farnham, 2011). Mice lacking *Kap1* die prior to gastrulation, indicating that KAP1 is a critical regulator of normal development and differentiation (Cammas et al., 2000). KAP1 interacts with histone methyltransferases and HDACs via a C-terminal PHD and bromodomain, and acts as a scaffold for various heterochromatin-inducing factors, such as heterochromatin protein 1 (HP1), the histone methyltransferase SETDB1 (also known as ESET), the nucleosome remodeling and histone deacetylation (NuRD) complex, and the nuclear receptor corepressor complex 1 (N-CoR1) (Quenneville et al., 2011, Schultz, 2002, Schultz et al., 2001). Thus it is not surprising that KAP1 mediates transcriptional repression

and establishes heterochromatin marks such as H3K9me3 through its binding of KRAB-ZFPs (Groner et al., 2010). Furthermore, via its interaction with KRAB-ZNFs and other transcription factors, KAP1 binds to thousands of sites in the human genome, including both 3'-coding exons of ZNF genes and promoter regions (Iyengar and Farnham, 2011). In addition, growing evidences demonstrated KAP1 could act as a robust transcriptional repressor when artificially recruited in multiple copies to promoters of reporter genes (Groner et al., 2010, Barde et al., 2009, Sripathy et al., 2006, Schultz, 2002, Ayyanathan, 2003). However, it is not clear if KAP1 endogenously regulates the transcript levels or epigenetic profiles of target genes. (Iyengar and Farnham, 2011)

In this chapter, I have shown that MBD4 directly interacts with KAP1 by reciprocal co-IP, and I have also identified the interacting domain of MBD4 responsible for the interaction, which further confirmed their robust association. Although my co-transfection of MBD4 and KAP1 in CMT93 cells is probably artificial, KAP1 indeed impacts the expression pattern and intensity of MBD4 in CMT93 cells, suggesting a mutual influence between them. Growing evidence support a link between DNA methylation and KAP1 together with the KRAB-ZFPs during the early embryonic period (Wiznerowicz et al., 2007). ZFP57 and its cofactor KAP1 associates with DNA methyltransferases and the hemimethylated DNA-binding protein UHRF1 (Quenneville et al., 2011, Zuo et al., 2012), and specifically binds to the H3K9me3-bearing, DNA-methylated allele of imprinting control regions (ICRs) in mouse ES cells. Additionally the methylated TGCCGC hexanucleotide was identified as the motif that is recognized by ZFP57 in all ICRs and in several tens of additional loci (Quenneville et al., 2011). I have shown that MBD4 interacts with UHRF1 and KAP1, in addition to our previous findings of its interaction with DNMTs. These MBD4 interacting proteins can selectively bind to unmethylated and hemi-methylated DNA, and have crucial roles in maintenance of H3K9 and DNA methylation at ICRs, supporting the key role of ZFP57 and KAP1 and their associating factors in regulation of imprinted genes during early embryogenesis. In addition, MBD4 possesses its MBD domain which can specifically target to the methylated DNA, its glycosylase part to repair the mismatches and involve in the possible demethylation activity, and most importantly its possible regulation ability to recruit co-factors to maintain a certain level of methyltransferases during and after the methylation activity. The association of MBD4 and KAP1 together with the DNMTs and UHRF1 suggest a possible role of MBD4 for the protection of specific loci against the demethylation, and/or dynamically regulate the methylation levels at the ICRs during the global demethylation stages. However, there may

be redundant proteins in this system, possibly other methyl-CpG binding proteins and another thymine glycosylase TDG.

Interestingly, in an assay using siRNA-mediated *Kap1* knockdown, It was shown that endogenous KAP1 negatively regulated TNF-α-induced IL-6 production in HeLa cells, and subsequently negative regulatory effects of KAP1 on TNF-α-induced NF-κB transcriptional activity by influencing the interactions among STAT3, p300, and NF-κB/p65 were identified. These data correlate with the possible biological significance of MBD4 in the inflammatory responses in rheumatoid arthritis discussed in chapter 5. Additionally the NF-κB associated gene transcripts are up-regulated in *Mbd4*$^{-/-}$ MEFs (chapter 5), supporting a possible mechanism, in which MBD4 may recruit transcriptional factors such as KAP1 to repress the transcription of NF-κB associating factors. In addition, KAP1 and UHRF1 have been implicated in systemic inflammatory disorders such as systemic lupus erythematosus (SLE), and the other potential MBD4 interacting partner TRIM27 from my screen has implicated in rheumatoid arthritis as well. Additionally both ACF1 and KAP1 have been found to associate with N-CoR1, the nuclear receptor corepressor complex, which may enable a repressive mechanism that MBD4 may be involved in. It is noteworthy that both ACF1 and KAP1 have the C-terminal PHD and BROMO domain, suggesting a possible common machinery (at least in part) to perform the transcriptional repression. KAP1 has also been implicated in tumor development. Increased KAP1 protein levels are found in liver, gastric, lung, breast, and prostate cancer, and gastric cancer patients and high levels of KAP1 have been linked to a significantly lower survival rate (Silva et al., 2006, Beer et al., 2002, Yokoe et al., 2009, Iyengar and Farnham, 2011). Most of these tumours specificity coincide with the ones MBD4 may associate with (Bader et al., 1999). RNAi knockdown of KAP1 in gastric cancer cells causes dysfunction in cell growth with an accumulation of cells in the G1 phase of the cell cycle (Wang et al., 2005, Iyengar and Farnham, 2011), and reduction of KAP1 stimulates p53 levels response to DNA damage, suggesting that KAP1 may be involved in a pathway to suppress apoptosis. Consistently, MBD4 interacts with FADD and DNMT1, and has the capability to induce apoptosis where the DNA damage is lethal and the mismatches cannot be repaired. This had led to a possible proposal that in additional to the putative role of transcriptional repression, the MBD4-Kap1 complex may play a critical role in proliferation and differentiation of both normal and tumour cells regarding the repair and apoptosis pathways.

PRMT5

Prmt5 (Protein arginine methyltransferase 5; also known as JBP1) is an type II arginine methyltransferase that has multiple cytoplasmic and nuclear substrates including histones H2A, H3 and H4, as well as transcription factors and co-regulators, and is involved in the regulation of the gene expression (Cesaro et al., 2009). Mice lacking *Prmt5* is early embryonic-lethal due to the abrogation of pluripotent cells in blastocysts (Tee et al., 2010). PRMT5 has multiple roles in germ cells, and as well as in pluripotency. Specifically, it has been shown that PRMT5 cooperates with the key regulator of unipotent primordial germ cells (PGCs) specification protein BLIMP1/PRDM1, to establish and propagate PGC from embryonic day 7.5 (E7.5) to E11.5 (Ancelin et al., 2006). In addition, PRMT5 is also important for the derivation and maintenance of pluripotent ES cells (Tee et al., 2010). It has been shown that PRMT5 is upregulated in the cytoplasm together with STAT3 during the derivation of ES cells, and associates with MEP50, contributing to the repression of differentiation genes by methylation of cytosolic histone H2A (H2AR3me2s) in ES cells (Tee et al., 2010). Moreover, PRMT5 is a key mediator component of the ZNF224 repressor complex, which interacts and depends on the corepressor KAP1 and the KRAB-A domain to perform the chromatic modification (Lupo et al., 2011). In the ZNF224 repressor complex, PRMT5 was demonstrated to methylate H4R3 on the nucleosomes surrounding the promoter region, which is associated with the repression of aldolase A gene expression (Lupo et al., 2011).

In this chapter, I have characterized the interaction between MBD4 and PRMT5 and KAP1, supporting a molecular model that MBD4 may recruit the corepressor KAP1 and the histone modifying enzymes such as PRMT5, to negatively regulate the expression of genes involved in the cell cycle. The formation of the ZNF224 transcriptional repression complex including KAP1 and PRMT5 at the human aldolase A gene as the transcriptional repressor suggests a possible role of MBD4 at these loci. The interacting domain of ZNF224 responsible for the association with PRMT5 was mapped between the second and third zinc finger motifs, which are conserved in several KRAB-ZFPs and are evolutionarily related. This suggests that the model where PRMT5 were centrally involved in together with its protein partners MBD4 and KAP1, may mediate the transcriptional repression of many other KRAB-ZFPs (Lupo et al., 2011). A recent study has also investigated the hypothesis whether Polycomb-mediated repression requires PRMT5 and proposed a model that PAX2/GRG4 recruit PRMT5 and promote symmetric H4R3 dimethylation (H4R3me2s), subsequently leading to PRC2 recruitment and H3K27me3 (Patel et al., 2011). It is very interesting that PRMT5 was linked to DNA methylation, where symmetric methylation of H4R3me2s by PRMT5 can recruit DNMT3A to the human β-globin locus, which is required for the DNA methylation

and silencing of this gene, and subsequently initiate the epigenetic silencing (Zhao et al., 2009). However, the structural basis of DNM3A does not support a direct binding to H4R3me2s peptide (Patel et al., 2011). My data of the direct interaction between MBD4 and PRMT5 provide a interesting link to support this model, suggesting an initiation of the most common marks H4R3me2s found in silent chromatin, followed by appropriate DNA methylation at designated loci.

Most interestingly, various chromatin-modifying activities associate with each other through the interplay between chromatin remodelers and protein arginine methyltransferases such as PRMT5 (Pal and Sif, 2007). PRMT5 is stimulated by the association with both BRG1 and BRM-based SWI/SNF, as well as associates with the mSIN3/HDAC corepressor complex, the components of the multisubunit nucleosome remodeling and deacetylase (NURD) complex, and symmetrically methylates MBD2 and negatively regulates its DNA binding activity (Guezennec et al., 2006, Tan and Nakielny, 2006, Pal and Sif, 2007). In particular, The biochemically and functionally distinct BRG1-based SWI/SNF complex can form into two separate complexes (BRG1 complex I and II), which differ by the presence of mSIN3/HDAC and PRMT5 corepressor proteins in BRG1 complex I (Pal and Sif, 2007). My data of the interaction between MBD4 and PRMT5 in addition to the previous reported MBD4 interaction with SIN3a/HDAC provide interesting link between the two BRG1 complexes I and II. Noticeably BRG1 interacts specifically with zinc finger containing transcription factors like KLF, Sp1, GATA, and type II hormone receptors like RAR and RXR (Kadam and Emerson, 2003), almost all of which are the putative targets that MBD4 may regulate (Chapter 5, by gene expression analysis of *Mbd4*$^{-/-}$ MEFs). This support a role of MBD4 to recruit chromatin remodelers such as BRG1 (SWI2/SNF2-related chromatin remodeling complexes) and ACF1 (ISWI-related chromatin remodeling complexes), and protein arginine methyltransferases such as PRMT5, subsequently alter chromatin structure to regulate and/or facilitate the transcriptional repression.

Type II arginine methyltransferases such as PRMT5 generate symmetric dimethylarginine, in contrast to the type I enzymes including PRMT1 and PRMT3 in the PRMT group that catalyze sequential formation of monomethylarginine and then asymmetric dimethylarginine (Aravind et al., 2011). PRMT5 and PRMT1–4 have highly conserved areas of sequence homology (Krause et al., 2007), thus it is not surprising that the tertiary structure of PRMT1 and PRMT3 are probably very similar to that of PRMT5. Consistently, PRMT1 and PRMT3 interact with MBD4 in my reciprocal IP that are simiar to PRMT5, albeit in my MBD4 interacting domain characteristics, PRMT1 cannot be IPed (preliminary data), which need further validation. It is noteworthy that the type I enzyme such as PRMT1 is also

implicated in diverse processes like transcription, protein localization, and signal transduction (Strahl et al., 2001). PRMT1 can directly associate with the p160 family of nuclear receptor coactivators and methylate histone H4R3 and facilitate transcription driven by the androgen receptor (Pal and Sif, 2007). Moreover, PRMT1 can also methylate N-terminal arginine residues in NFAT interacting nuclear protein 45 (NIP45) to positively regulate cytokine production from Th cells (Mowen et al., 2004). In addition, PRMT1 can not only methylate and increase the DNA binding activity of the nuclear receptor HNF4, but also be recruited by p160 coactivators to HNF4 target genes results in H4R3 methylation and transcriptional activation (Barrero and Malik, 2006, Pal and Sif, 2007). It is very interesting that the cluster of the orphan nuclear receptor HNF4 and its cofactor and target genes were at the top ranking of both MBD4 protein interaction networks (Chapter 3) and *Mbd4*$^{-/-}$ gene expression changes (Chapter 5), suggesting a possible role of MBD4 in transcriptional activation at the orphan nuclear receptors such as HNF4 cluster and their specific pathways. ZNF224 was demonstrated to act as a transcriptional co-activator of WT1 on VDR promoter (Lupo et al., 2011). Consistently a decreased expression of VDR was found in the human erythroleukemia cell line K562 when ZNF224 was depleted in these cells (Florio et al., 2010), suggesting a role for ZNF224/WT1 interaction in this pathway through the transcriptional activation of VDR (Lupo et al., 2011). The interaction between MBD4 and PRMT5, the component of ZNF224, suggesting a role of MBD4 in the transcriptional activation of VDR through the mediation of the ZNF224 complex, possibly through the putative active DNA demethylation pathway that I have discussed, or a possible mechanism through an apoptosis pathway. In support to the latter, MBD4 has been shown to act as a sensor to induce apoptosis (Screaton et al., 2003, Ruzov et al., 2009), while vitamin D has been shown to induce apoptosis in leukemia cells (Kizildag et al., 2009), and the complex of ZNF224 and the cancer testis antigen DEPDC1 was identified to repress the transcription of the A20 gene, which is a negative regulator of the NF-κB-mediated antiapoptotic pathway (Lupo et al., 2011, Florio et al., 2010). Thus, it is very interesting to investigate the role of MBD4 in the transcriptional derepression at the type II non-steroid hormone receptors including adopted orphan nuclear receptors (RXR)/the RXR heterodimer receptors, and/or orphan receptors such as HNF4a, in the context of the pathways of putative active DNA demethylation and/or apoptosis.

Last but not least, PRMT5 was identified as JAK2 interacting protein in a yeast two-hybrid experiment (Pollack et al., 1999). *JAK2* tyrosine kinase mutation *JAK2-V617F* is present in about 95% of polycythemia vera patients and about 50% of essential thrombocythemia and primary myelofibrosis patients (Tefferi et al., 2009, Cross, 2011), the

two major forms of myeloproliferative neoplasms (MPN). Mouse models that express the MPN-associated *JAK2* mutants recapitulate the human MPN-like disorders, confirming these *JAK2* mutations indeed pay a key role in the morphogenesis of these disorders (Cross, 2011). Interestingly a recent research demonstrated PRMT5 can be phosphorylated by the MPN-associated JAK2 mutants, and subsequently inhibit the histone methyltransferase activity of PRMT5 (Liu et al., 2011), suggesting a link between the JAK/STAT pathway and transcriptional regulation through epigenetic modifications in MPN formation. My finding of the interaction between MBD4 and PRMT5 suggests a possible role of MBD4 in these MPN diseases through impairment of signaling with PRMT5 and/or its putative role in the possible global DNA demethylation during mouse erythropoiesis (Shearstone et al., 2011).

4.3.3. Transcriptional repression mediated by MBD4: A role of MBD4 that might be underestimated?

Methylation at 5 position of cytosine is the most common modification in higher eukaryotes genomes. In mammals, the methylation modification occurs almost exclusively in CpG dinucleotides (Ehrlich and Wang, 1981, Riggs and Jones, 1983, Doerfler, 1983). Methylated CpG dinucleotides are ~70% of total mammalian CpG, while the ~1% CG-rich regions called CpG islands are usually unmethylated, which mostly coincide with the promoter of protein-coding genes (Fraga et al., 2003). Methylation of CpG dinucleotides at promoter region is generally associated with transcriptional repression (Antequera et al., 1990, Jaenisch and Bird, 2003). In this context, the methyl-CpG binding proteins (MeCPs) that are capable of recognition of methylated CpG dinucleotides are proposed to play a central role in DNA methylation associated transcriptional repression (Ballestar and Wolffe, 2001). The MeCPs contain ~70-75 amino acids methyl-CpG binding domain (MBD), conserved between 50-70% similarity (Hendrich and Bird, 1998). As introduced in chapter 1, The MBD domains of MeCP2, MBD1, MBD2, and MBD4 are necessary and sufficient for DNA binding, and recognise and bind to single symmetrically methylated CpG dinucleotides in vitro (Nan et al., 1993, Hendrich and Bird, 1998, Hendrich et al., 1999). Repression by MeCPs can be conducted by recruitment of their associated co-repressors including histone deacetylases, the Mi-2 or NuRD complex, and mSin3 corepressor (Ng et al., 2000, Yu et al., 2000, Zhang et al., 1999, Nan et al., 1998, Ng et al., 1999, Kaludov and Wolffe, 2000).

MBD4 is the only protein that possesses a glycosylase repair function within MECPs, in addition to its N-terminal MBD domain. MBD4 was reported to associate with

the hypermethylated promoters of tumor suppressor and DNA repair genes including p16INK4a and hMLH1 genes, and recruit co-repressors such as Sin3A, HDAC1, and RFP (Fukushige et al., 2006, Kondo et al., 2005). However, globally the phenotype of *Mbd4* knockout mice was very mild, with the uplifted thymine mutation rate throughout genome possibly due to the lack of the thymine glycosylase repair activity in the knockouts (Millar et al., 2002). Mice homozygous for *Mbd4*$^{-/-}$ were viable and fertile, arguing against a major transcriptional repressive role of MBD4 in early development and reproductive system. In addition, this also suggests the transcriptional repression that MBD4 may mediate is unlikely the general or global repression. However, this does not exclude the possible important role of MBD4 in differentiated cells and/or late developmental stage to repress the specific genes or a specific set of genes, which shall not expressed in the specialized states. The transcriptional repression mechanisms may be placed but not limited into three categories, including inhibition of the basal transcription machinery, ablation of activator function, and remodeling of chromatin (Gaston and Jayaraman, 2003). In my study, I identified a cohort of MBD4 interacting proteins including KAP1, PRMT5, UHRF1, and USP7 (Figure. 4.18) that in molecular detail potentially orchestrate transcriptional repression via some of these cross-talking mechanisms. Although the validated protein interactions of MBD4 are complicated in functional diversity, a common theme of these combinations of mechanisms that MBD4 and its interacting proteins may act at is most likely the methylated promoter dependent. In my working model, MBD4 may recognise and bind to 5mCpGs at promoter sites of specific genes, and recruit its association proteins to induce/maintain a repressive chromatin state at downstream nucleosome (Figure. 4.18). On one hand, these include histone modifying complexes (HMC) such as PRMT5, UHRF1, and USP7, which has been reported to directly modify the status of ubiquitylation (Khoronenkova et al., 2011, Unoki et al., 2009) and arginine methylation (Cesaro et al., 2009) of histones in chromatin remodelling (Figure. 4.18. upper side; A). On the other hand, KAP1 has been demonstrated as a scaffold protein to assemble epigenetic machinery (Iyengar and Farnham, 2011), and form repressive complexes with KRAB-ZFPs such as ZFP53 during the early embryonic period (Riclet et al., 2009). In addition, PRMT5 has been found as a key mediator component of the ZNF224 repressor complex, which interacts and depends on the corepressor KAP1 and the KRAB-A domain to perform the chromatic modification (Lupo et al., 2011)(Figure. 4.18. lower side; B). All of these complexes may be recruited by MBD4 in the context of specific tissue types and developmental stages, and gene-specific repressor proteins often act in a ‘short-range’ binding proximal to the promoter, or a ‘long-range’ at a distance from the promoter sites (Mannervik, 1999). The MBD domain of Mbd4 can specifically recognise methylation

marks at single CpG level, most of which overlap with promoter sites in the genome. In addition, in chapter 6 I have shown this methyl-CpG recognition ability of MBD4 is well conserved during the vertebrate evolution, despite of the poor conservation of its primary sequences. These support a general theme that MBD4 may act as a repressor mainly at the binding sites proximal to promoters of a set of specific genes. However, the possibility of 'long-range' repression that MBD4 might involve in is not exclusive, as the two mechanisms are often cooperative depending on the context of its binding sites and its exact mode of action (Gaston and Jayaraman, 2003, Courey and Jia, 2001).

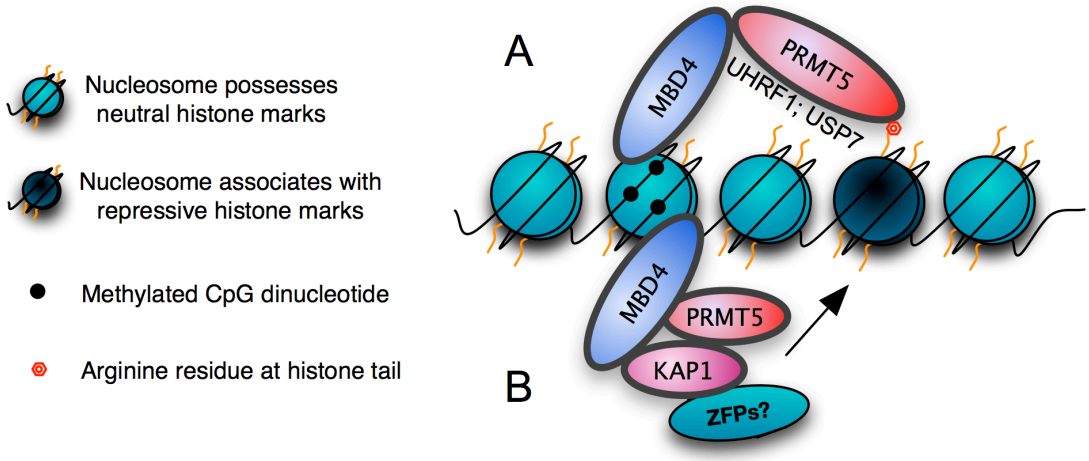


Figure. 4.18. A working model for transcriptional repression of MBD4 by recruiting histone modifying complexes (HMC) such as PRMT5, UHRF1, and USP7 (Upper side; A), and corepressor complex including KAP1, PRMT5, and specific zinc finger proteins (lower side; B). 5mCpG (Black dots) can be recognised and bound by MBD4, which can recruit HMC and corepressors to induce/maintain a repressive chromatin state (dark nucleosome).

4.3.4 Concluding Remarks

In sum, I have validated a number of MBD4 association candidates by reciprocal co-IP and IF colocalisation. The significant MBD4 associations with UHRF1 and USP7 suggest that MBD4 might have a role in the maintenance of DNA methylation through its direct interaction with DNMT1 (previously shown)/UHRF1 (novel) /USP7 (novel) possibly by regulating the stability of DNMT1 through UHRF1 and USP7. In addition, MBD4 might have cross-talk to histone modification and chromatin remodelling to act as transcriptional repressor through partners such as KAP1, PRMT5 and/or ACF1. In general positive

validation by IP and IF demonstrates the robustness of the initial screens, and implies that MBD4 may impact several transcriptional and epigenetic networks along with a number of nuclear pathways that include transcriptional repression, repair survey, and regulation of DNA methylation machinery. In addition, the interaction between MBD4 and CFP1 and SETD1a may be true, as they were validated by one-side FLAG trap (Table 4.1&Figure. 4.1 B lower) as well as partially overlap with MBD4 expression in CMT93 cells. CFP1 not only interact with but also dramatically influence the stability of exogenously expressed MBD4 in human HEK293T cells (Figure. 4.15 A&B), suggesting there might be a role of MBD4 in transcriptional activation by its association with CFP1 and SETD1a, which needs further investigation. Lastly, although the MBD4 interaction with SC35 was not validated by co-IP, SC35 did indeed inhibited the expression of exogenous MBD4 in HEK293T cells (Figure. 4.15 C), suggesting there is certain connection between MBD4 and splicing factor SC35, which need further validation.

However, It is noteworthy that there are a number of limitations of the experiments. Firstly, all the assays in this chapter were exogenous and may be artificial. This needs particular attention in the cases of KAP1 and PRMT5, which showed unusual condensed distribution in cytoplasm, and their co-localisations with MBD4 does not exist, despite of their robust co-IP results. Additionally the MBD4 associations with ACF1 need to be further validated as its immunoprecipitation was only successfully performed once, and if overexpressed ACF1 could resemble its endogenous distribution is unknown. Secondly, the IF could show the localisations between MBD4 and its association proteins at a 2D manner, but if they are in the same location is unknown, which requires confirmation by confocal microscopy. Thirdly, some MBD4 association proteins could affect the expression intensity and pattern of GFP-MBD4 in 293T cells. However, if MBD4 could influence them in the same cell system is unknown. Fourthly, I have identified the MBD4 domains responsible for its associations with several validated proteins. However, further investigations of the domains of these proteins that interact with MBD4 are needed for a better understanding of their associations. Finally, I could show the associations of MBD4 and KAP1, PRMT5, USP7, UHRF1 and ACF1. However, the biological relevance of these associations is unknown. Additional experiments are required. For example, luciferase based reporter assays can be performed to show if MBD4 indeed act as a transcriptional repressor, with corepressors such as KAP1, PRMT5. Another example would be using cell system like *p53*$^{-/-}$ MEFs to investigate what changes would happen when MBD4 expresses with USP7 or UHRF1, in the context of apoptosis and DNA damage responses. It was a very difficult question regarding the biological significance of MBD4, partially because the phenotype of

Mbd4 knockout mice is very mild, and other MeCPs and/or glycosylase protein might compensate its role in the system. However, the large number of MBD4 associations in a number of significant pathways especially in transcriptional repression and regulation of methylation suggest the role of MBD4 in the system may be indispensable. Further investigations are required, particularly in terms of the gene expression changes when *Mbd4* is inactivated, and/or in specific processes such as haematopoiesis and specific pathways such as the possible active DNA demethylations.

Chapter 5: Profiling of transcriptional aberration in two independent cell lines that lack MBD4 activity

5.1. Introduction

In Chapter 3 and 4, I have identified and validated a number of MBD4 association proteins, suggesting that MBD4 may impact several significant pathways, particularly in terms of transcriptional repression, repair survey, and regulation of DNA methylation machinery. However, mice lacking MBD4 are both viable and fertile, arguing that the role of MBD4 might be dispensable or compensatable (Millar et al., 2002). This might be because there are a number of MeCPs and/or glycosylases in the system with functional and domain structural similarity, suggesting their redundant roles may compensate the loss of function of individual proteins. For example, almost all the MeCPs have been associated with transcriptional repression (Ng et al., 2000, Yu et al., 2000, Zhang et al., 1999, Nan et al., 1998, Ng et al., 1999, Kaludov and Wolffe, 2000), which are particularly important for stable silencing the genes that are not expressed in certain types of cells. In addition, in mammals MBD4 and TDG are the only two proteins that have thymine glycosylase activity at least in vitro (Jacobs and Schär, 2011). *Mbd4*$^{-/-}$ mice indeed showed ~3 fold enhanced CpG mutability but this mutation rate is far less than that resulting from the removal of all the thymine repair machinery (Millar et al., 2002). Consistently, genome-wide C to T mutagenesis has not been observed so far in either MBD4 or TDG-deficient cells (Cortázar et al., 2011, Jacobs and Schär, 2011), suggesting MBD4 and TDG might compensate for each other at least in part for thymine mismatch repair. In addition, both MBD4 and TDG have been reported to associate with transcriptional repression (Fukushige et al., 2006, Kondo et al., 2005a, Cortázar et al., 2011), and interact with DNA methyltransferases (Kim et al., 2009b, Métivier et al., 2008, Ruzov et al., 2009), supporting the hypothesis that they might be complementary in terms of these important processes in the mammalian system. Although catalytic inactivation of TDG in mice is embryonic lethal (Cortázar et al., 2011), this suggests TDG might have a developmental role for which MBD4 cannot compensate in early developmental stages. However, this could not exclude their compensatory roles in post-embryonic stage, especially in particular types of tissues. In particular, the mild phenotype of MBD4 knockout mice supports that TDG might have the ability to compensate at least in part for MBD4 inactivation.

Moreover, although *Mbd4*$^{-/-}$ mice exhibit a very mild phenotype under normal conditions, this does not preclude the possibility that the loss of MBD4 function would be significant where the system were exposed under stress such as oxidative factors, ageing, and/or pro-inflammatory environment. Experimental mice are usually sacrificed within a few months, but the onsets of some subtypes of complex diseases in cancer, autoimmune disease,

and ageing related diseases are usually after several decades, thus the full effects of *Mbd4* deficiency may not have been recognised. Interestingly, inactivation of *Mbd4* was reported to associate with increased tumour formation, but only when combined with *Apc* min oncogenic background in mice (Millar et al., 2002). In addition, the *Mbd4* mutations are associated with several specific cancers and inflammatory disorders (Bader et al., 1999, Riccio et al., 1999, Abdel-Rahman et al., 2008, Saito et al., 2001, Xi et al., 2008), suggesting a potential value of investigation of MBD4 in understanding the pathogenesis of these diseases. However, the potential functional roles of MBD4 remain largely unknown.

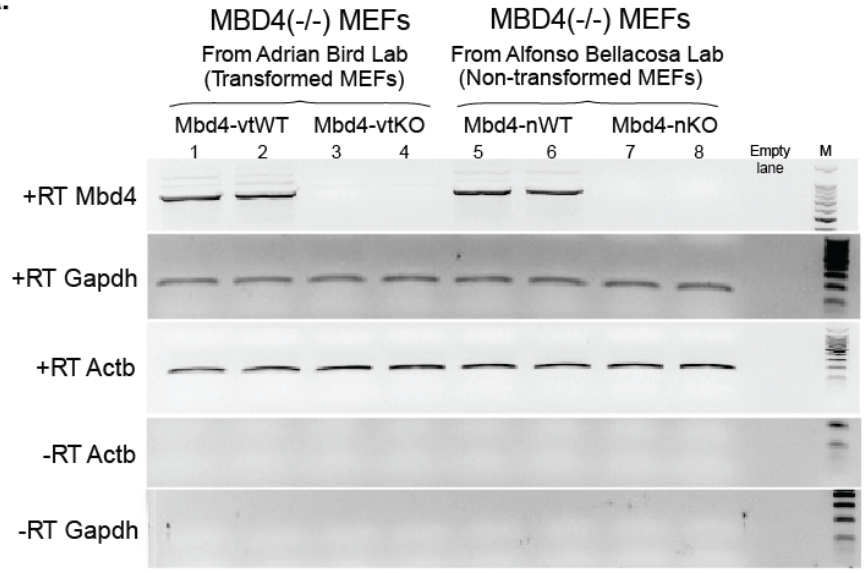
The independent labs of Adrain Bird and Alfonso Bellacosa generated *Mbd4*-deficient mice (Cortellino et al., 2003, Millar et al., 2002). This has enabled me to analyse the genome-wide gene expression changes caused by MBD4 activity using mouse embryonic fibroblasts (MEFs) generated from these independent *Mbd4*$^{-/-}$ mice. Although the mild phenotypic effects of these *Mbd4*$^{-/-}$ mice have been well characterised, their changes at gene expression level are unclear. In this chapter, I first set out to investigate the most significant genes that were mis-regulated by lack of MBD4 activity. Following a clue that the expressions of a cohort of insulin-like growth factor-binding proteins were dramatically changed, I tested the possibility that MBD4 could directly interact with estrogen receptor α (ERα). With the long-term aim of finding possible biological significance of MBD4, I did pathway analysis of MBD4 in parallel with TDG, to investigate a possible role of MBD4 in pro-inflammatory responses in acquired conditions such as cancer and autoimmune disease. In response to this unexpected relationship, I focussed on the role of MBD4 in transcriptional regulation and identified several potential MBD4 targets including developmental genes *Sox9*, *Klf2*, *Klf4*, supporting a hypothesis that MBD4 might be responsible for stable expression of these developmental genes in specific types of tissues, which needs further investigation in the future.

5.2. Results

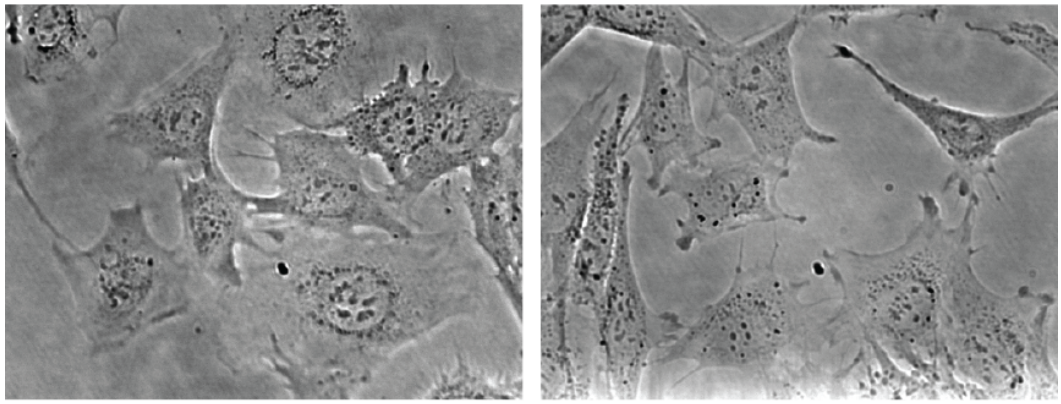
5.2.1. Identification of gene candidates significantly mis-regulated in two independent Mbd4$^{-/-}$ *MEFs*

The two *Mbd4*$^{-/-}$ cell lines were the virus-transformed cell line (labelled vtWT and vtKO) from the Adrian Bird Lab (Millar et al., 2002) and the non-transformed cell line (labelled nWT and nKO) from the Alfonso Bellacosa Lab (Bellacosa et al., 1999) (Figure. 5.1).

A.



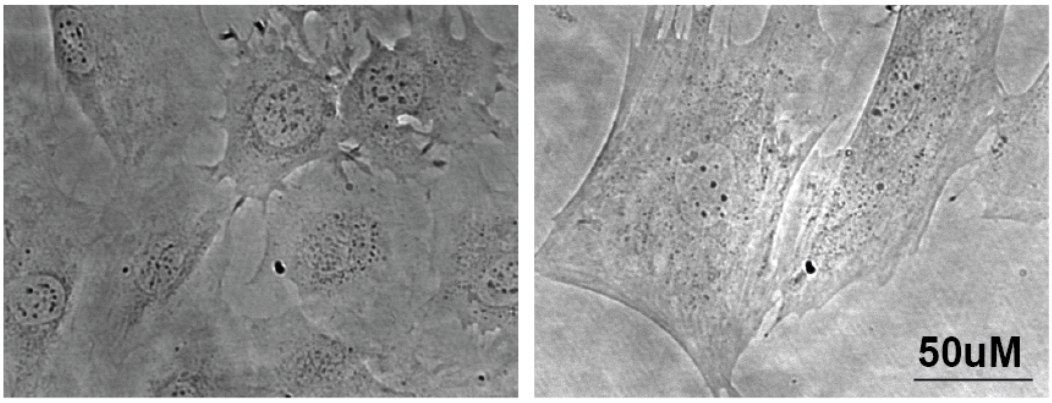
B.



Mbd4 transformed WT
(Mbd4 vtWT)

Mbd4 transformed KO
(Mbd4 vtKO)

C.



Mbd4 non-transformed WT
(Mbd4 nWT)

Mbd4 non-transformed KO
(Mbd4 nKO)

Figure. 5.1. Confirmation of MBD4 expression in *Mbd4*$^{+/+}$ and *Mbd4*$^{-/-}$ in the two independent cell lines for Illumina's whole-genome expression mouse WG-6 BeadChip. (A) RT-PCR analysis of *Mbd4* transcript expression in the two independent cell lines. *Actb* and *Gapdh* transcript expression are shown as positive controls. RT indicates the reverse transcription reaction. The morphology of *Mbd4*$^{+/+}$ and *Mbd4*$^{-/-}$ MEFs were visualised by light microscopy. (B) The virus transformed MEFs (vtWT & vtKO). (C) The non-transformed MEFs (nWT & nKO).

cRNA was prepared from total RNA extracted from four pairs of vtWT and vtKO transformed cell lines, and two pairs of nWT and nKO non-transformed cell lines. Both cell lines were cultured at least five days from the frozen stocks and split twice before harvesting using Trizol reagent. RT-PCR confirmed there were no wild type contamination and no expression of MBD4 in both knockout cell lines (Figure. 5.1 A). The cRNAs were then hybridized to the Mouse WG-6 v2.0 Expression BeadChip (Illumina, San Diego, CA). Differentially expressed genes were detected using the Illumina GenomeStudio Gene Expression Module (see methods). In general, the microarray analysis followed normalization of the quantile or average normalisation method, and based on gene-profiling (sum of several probes representing one gene) or probe-profiling, because the Illumina mouse WG6 system has ~45281 probe sets to represent ~19100 genes.

Initially, most stringent criterion to identify the most significant candidates present in both cell lines was adopted. This criterion was that their mis-regulated expression alterations must be at least five fold in both cell lines and consistently overlap between both cell lines. In addition, their significant expression changes must be identified from either gene-profiling or probe-profiling approaches as well as from the profiling of both quantile and average normalisations. Seventeen transcripts were identified, including seven genes from both gene- and probe- profiling, and another ten genes from probe-profiling only (Table. 5.1).

Top Candidates: highly significant and confident in BOTH gene AND probe ranking			
Igfbp4	20 fold	downregulated	insulin-like growth factor binding protein 4
Sdpr	13 fold	upregulated	serum deprivation response
Wnt4	10 fold	downregulated	wingless-type MMTV integration site family, member 4
Dio3	9 fold	downregulated	deiodinase, iodothyronine, type III
Gfp1	7 fold	downregulated	glutamine--fructose-6-phosphate transaminase
Odz4	7 fold	downregulated	odz, odd Oz/ten-m homolog 4 (Drosophila)
Cryl1	6 fold	upregulated	crystallin, lambda 1
High Confidence Candidates: highly significant and confident in probe ranking only			
Mtap2	41 fold	upregulated	microtubule-associated protein 2
Ifitm1	27 fold	downregulated	interferon induced transmembrane protein 1
Tox	24 fold	upregulated	thymocyte selection-associated high mobility group box
Gjb2	23 fold	upregulated	gap junction protein, beta 2, 26kDa
Cp	12 fold	downregulated	ceruloplasmin (ferroxidase)
Igfbp5	11 fold	upregulated	insulin-like growth factor binding protein 5
Anxa4	10 fold	upregulated	annexin A4
Emb	7 fold	downregulated	embigin
Pcsk6	6 fold	upregulated	proprotein convertase subtilisin/kexin type 6
Il17d	6 fold	upregulated	interleukin 17D

Table. 5.1. Highly significant and confident genes mis-regulated in both two independent *Mbd4*$^{-/-}$ MEF cell lines. All of these candidates are at least five-fold up- or down- regulated, consistent in both cell lines and analyses of quantile and average normalisation.

Quantitative reverse transcription PCR (qRT-PCR) was used to validate alterations in gene expression detected in *Mbd4*$^{-/-}$ MEFs using genome-wide microarray analysis. qRT-PCR primer pairs were designed as an intron-cross manner. Firstly I tested two down-regulated (*Igfbp4* & *Dio3*) and two up-regulated candidates (*Igfbp5* & *Cry1*) from the most significant candidate pool (seventeen candidates, see table 5.1). The cDNAs used for this qRT-PCR validation was the same as the ones that were put on the beadarray. The qRT-PCR results from all these genes were consistent with the microarray findings in both cell lines (Figure. 5.2 A). Additionally another four candidates from probe-profiling based analysis within the same candidate pool were tested in a batch of re-cultured cDNA samples. This time only vtKO cell line was used because the nKO cell line can only survive for limited passages, as they have not been virus-transfected for immortal growth. Additional qRT-PCR confirmed the microarray findings, with fold change slightly lower than the microarray results. Most of them (*Lfitm1*, *Wnt4*, and *Cp*) were still more than five fold changed (Figure. 5.2 B). Only one candidate *Anxa4* was lower than five-fold change (about two fold), but it was still consistent and significant (Figure. 5.2 B right). Altogether, the validation of

microarray results showed the candidates identified by Illumina beadarrays in two independent knockout cell lines were robust.

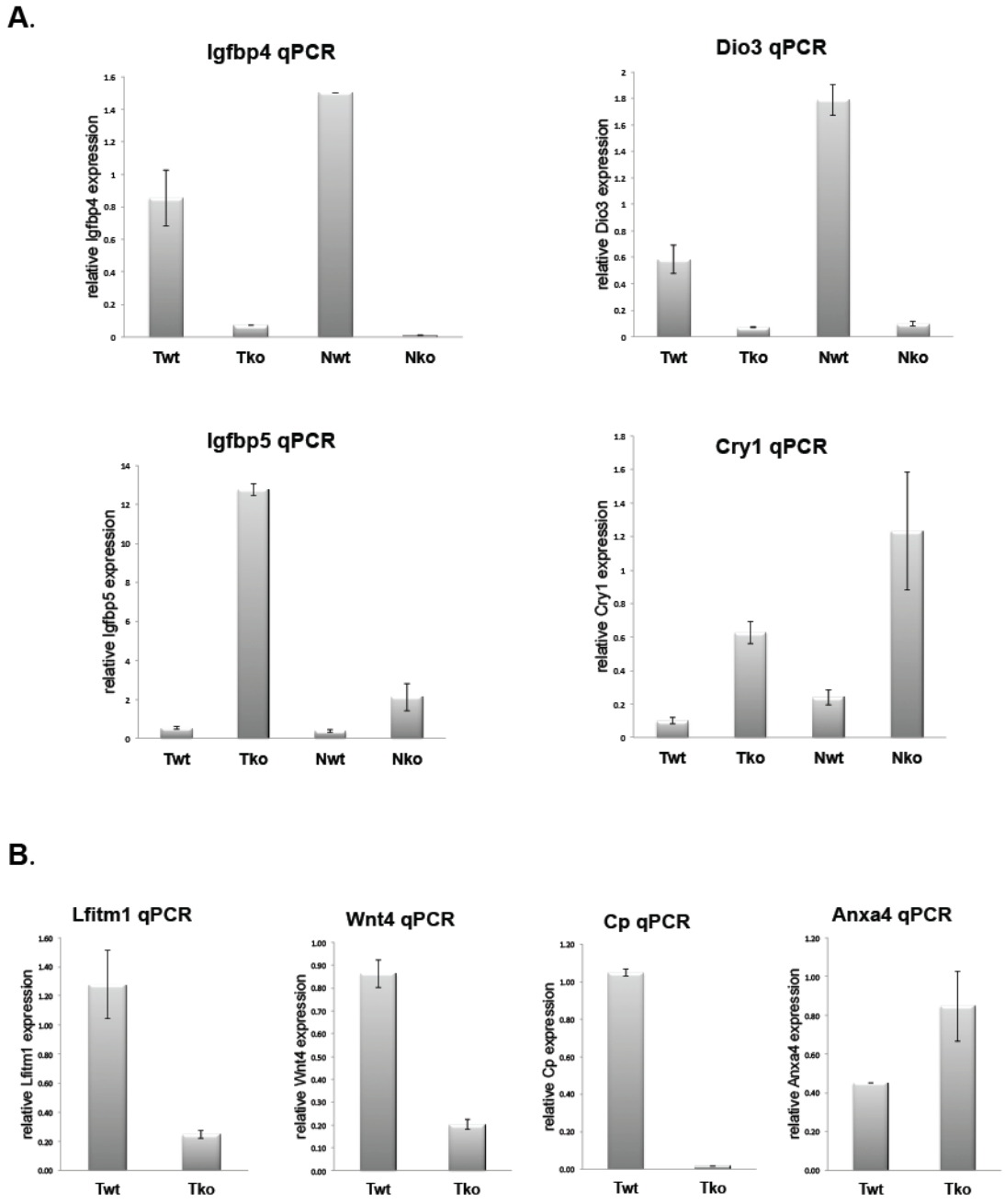


Figure. 5.2. qRT-PCR Validation of expression microarray results. A. Quantitative RT-PCR validation of four significant genes highly mis-expressed in *Mbd4*$^{-/-}$ MEFs in both cell lines. The relative expression of each genes were normalised to *Gapdh*. B. qRT-PCR confirmation of another four significant genes highly mis-expressed in transformed *Mbd4*$^{-/-}$ MEFs. The relative expression of each genes were normalised to *Actb*.

5.2.2. MBD4 might potentiate transcription of estrogen regulated genes via a direct interaction with ERα

The high-stringent profiling strategy provided me a pool of seventeen robust candidates that are mis-regulated by lacking MBD4 activity. Two genes, *Igfbp4* and *Igfbp5* from the high-stringent candidate pool, and two another genes *Dusp4* and *Runx1* at lower fold changes (not in the pool but consistent in two cell lines with more than two fold changed) overlap with genes that are mis-regulated in response to tamoxifen exposure in an ER-positive ZR-75-1 xenograft model (Taylor et al., 2010). In response to this observation I co-transfected GFP- or FLAG- tagged full length MBD4 with an ERα expression vector in HEK293T cells, and identified a direct interaction between MBD4 and Estrogen receptor α (ERα) by co-IP (Figure. 5.3 A-B). Additionally, partial co-localisation of exogenous MBD4 and ERα was observed in murine CMT93 cells by IF microscopy (Figure. 5.3 C-D). Altogether these observations suggest a possible direct interaction between MBD4 and the Estrogen receptor α (ERα).

5.2.3. Two paternally imprinted genes were significantly mis-regulated in the two Mbd4$^{-/-}$ cell lines

From the candidate pool I also observed that a paternally imprinted gene, *Dio3* was dramatically down-regulated in the two *Mbd4*$^{-/-}$ cells lines. Therefore, all the imprinted genes in my microarray data were examined. Another paternally imprinted gene *H19* was also dramatically up-regulated in both cell lines (eight fold in vtKO, three fold in nKO). The other one at the same imprinting cluster as *Igf2* showed significant change, but its expression changes were not consistent across the two cell lines (~twelve fold upregulated in vtKO, ~twenty one fold down-regulated in nKO). In contrast, all the maternally imprinted genes are either no significant changes or undetectable from the beadarrays. The discussion and a working model regarding the transcriptional regulation of these paternally imprinted genes through MBD4 is in general discussion chapter.

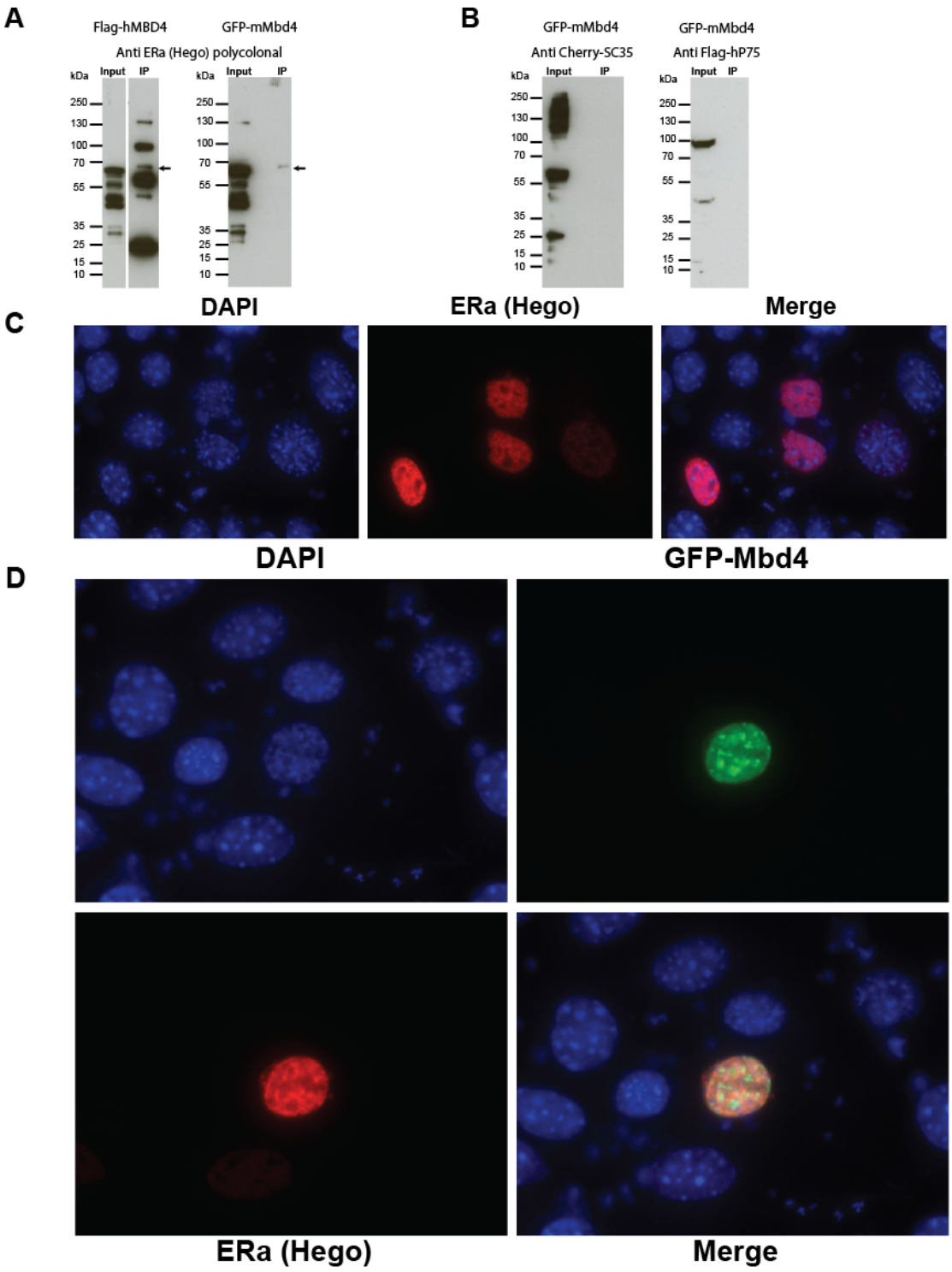

A
Flag-hMBD4
GFP-mMbd4
Anti ERa (Hego) polycolonal
kDa
Input
IP
250
130
100
70
55
35
25
15
10
B
GFP-mMbd4
Anti Cherry-SC35
GFP-mMbd4
Anti Flag-hP75
C
DAPI
ERa (Hego)
Merge
D
DAPI
GFP-Mbd4
ERa (Hego)
Merge

Figure. 5.3. Potential interaction between MBD4 and Estrogen receptor α (ERα) identified by co-IP and IF co-localisation. (A)&(B). Immunoprecipitation using 293T cells expressing GFP- or FLAG- MBD4 and a vector containing ERα (HEGO vector, See method), were lysed and IPed with an antibody directed against ERα. Association between MBD4 and ERα was analysed by sequential western blotting using indicated antibodies. (A) MBD4 confirmed to associate with ERα; Arrows indicate the ERα bands at the correct size. (B) Negative controls using the same GFP trap. The IP lanes are completely blank and negative. (C)&(D). ERα shows a diffused distribution in nuclear, partially overlaps with MBD4 and bright DAPI spots. (C) Distribution of ERα was compared with DAPI counterstaining. Vectors containing ERα were transiently transfected into mouse CMT93 cells, and were detected with anti-ERα antibody and Texas Red-conjugated secondary antibody. The overexpressed ERα exhibits diffused general distribution in nuclear. The bright DAPI spots were partially overlap with such diffused distribution of ERα. (D) GFP-mMbd4 and ERα were transiently expressed together and compared with DAPI staining. Overexpressed MBD4 colocalises with DAPI bright spots that are associated with heterochromatic chromosomes, and ERα shows diffused distribution in nuclei, and partially overlap with DAPI and MBD4 bright spots.

5.2.4. Identification of a clue pointing to possible biological significance of MBD4 and TDG in pro-inflammatory responses, and in acquired diseases such as cancer and autoimmune disease

The most stringent criterion enabled identification of a cohort of genes with the highest fold changes in both cell lines, which are most likely regulated by MBD4 activity; however, it discarded the clues at the systemic level, thus it is hard to predict the pathways and functions that MBD4 may be involved in. I then used a lesser, but still significant, stringency of two-fold expression change for MBD4 virus transformed vtKO cell line ($p<0.05$ according to t-test, n = 4), and three-fold expression change for MBD4 non-transformed nKO cell line ($p<0.05$ according to Mann-Whitney, n = 2). The reason of adopting higher threshold on *Mbd4* nKO is that only two pairs of samples were microarrayed. Thus a higher threshold could compensate the shortage of limited sample numbers performed and increase the robustness of the *Mbd4* nKO results. Microarray analysis of whole genome expression in *Mbd4*$^{-/-}$ MEFs compared to WT identified 639 differentially expressed transcripts in *Mbd4* vtKO, as well as 583 in *Mbd4* nKO (Figure. 5.4 middle). These lists are referred to the Mbd4vtKO and Mbd4nKO lists and can be found in full in the supplementary table 5.1 & 5.2. The numbers for up- or down- regulated genes in both cell lines are similar (Figure 5.4 left & right). In addition, I noticed a cohort of these perturbed transcripts overlaps between

both cell lines (137 transcripts in total, 75 consistent in both arrays, 32 up- and 43 down-regulated) (Figure 5.4 middle).

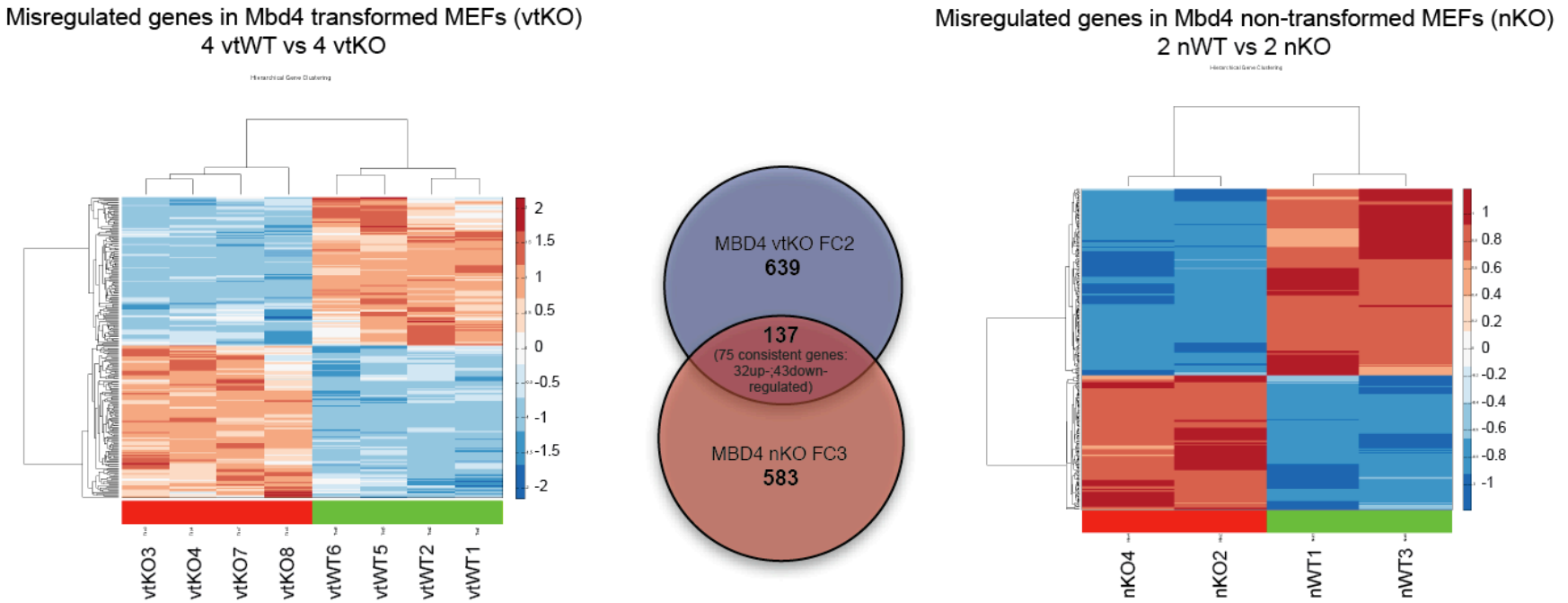


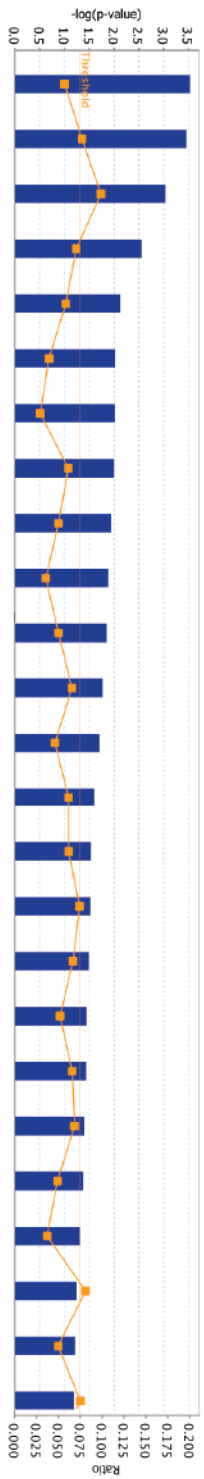
Figure. 5.4. Venn diagram and heatmaps of differentially expressed genes in the two independent *Mbd4*$^{-/-}$ MEFs. The names of the cell lines, number of up- and down- regulated genes, and the fold change threshold (FC) adopted are indicated. In heatmaps, Left: *Mbd4* transformed KO MEFs; Right: *Mbd4* non-transformed KO MEFs. The colour bars on the right indicate the log values of differential expression for the responsive colour intensities.

In order to identify the possible canonical pathways and biological functions that are responsive to MBD4 activity, data from the *Mbd4* vtKO and nKO gene lists was analysed through Ingenuity Pathway Analysis (IPA) system (Ingenuity Systems, Mountain View, CA, USA, www.ingenuity.com). The datasets containing differently expressed genes (FDR adjusted P<0.05 and Fold change (FC) >2 for vtKO & FC>3 for nKO) with their values of fold change were uploaded to the application (see methods). A number of canonical pathways including the functions of several type II hormone receptors (RXR, VDR etc.) were implicated with significant cluster values in both two independent $Mbd4^{-/-}$ cell lines (Figure. 5.5. bold and Green colour pathway categories). These include the previous reported MBD4 association with Vitamin D receptors (Kim et al., 2011), suggesting the pathways implicated are biologically meaningful.

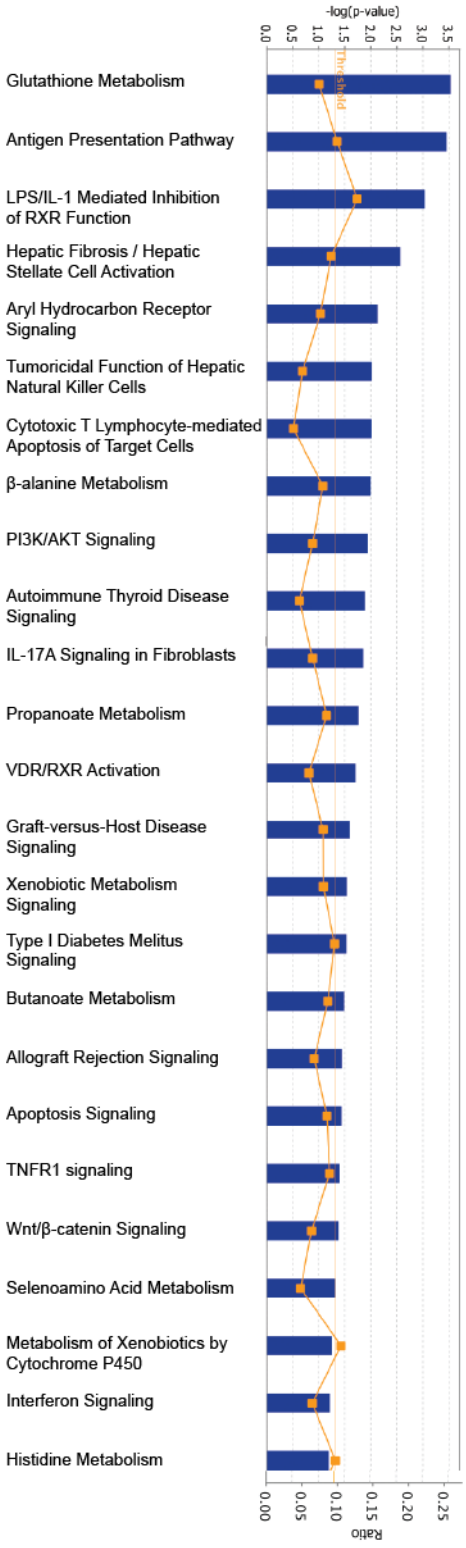
Unexpectedly, several other top pathways imply a clue for possible significance of MBD4 in conditions such as Rheumatoid Arthritis (RA) and hepatic fibrosis (Figure. 5.5 pathways 1, 8, 11 in vtKO line). However, most of the significance of these pathways was not consistent in another nKO cell line. This may be because nKO cells have high susceptibility to undergoing apoptosis, dying within limited numbers of passage (Figure 5.1 C right). I then ask if the only other mammalian thymine DNA glycosylase (Tdg) could also be implicated in these pathways, as there might be redundant function between MBD4 and TDG in the system as discussed above. The microarray data of whole genome expression in $Tdg^{-/-}$ MEFs compared to $Tdg^{+/-}$ is available at NCBI Gene Expression Omnibus under accession number GSE20693 (Cortázar et al., 2011). These $Tdg^{-/-}$ MEFs are virus transformed cell lines generated from the same lab of the *Mbd4* vtKO cell line, and microarrays were performed on the same Illumina mouse WG6 platform I used for $Mbd4^{-/-}$ cells lines, thus they are suitable for parallel comparison with my *Mbd4* pathway analysis. At a log fold change 1.2 ($p<0.05$ according to T test, n = 3), 180 differentially expressed transcripts of *Tdg* vtKO MEFs were uploaded to the IPA application, and the involved pathways of *Tdg* were compared with those of *Mbd4* in Figure. 5.5. Surprisingly a number of significant canonical pathways were implicated in both *Mbd4* vtKO and *Tdg* vtKO cell lines, all of which suggest a possible biological convergence between MBD4 and TDG in pro-inflammatory responses in acquired conditions such as cancer and autoimmune disease such as Rheumatoid Arthritis, and hepatic fibrosis (Figure. 5.5. bold and Red colour pathway categories). Specifically, 8 overlapping canonical pathways were implicated in both *Mbd4* and *Tdg* vtKO profiling: 1. Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis; 2. Human Embryonic Stem Cell Pluripotency; 3. Hepatic Fibrosis /

A. Mbd4 vtKO profiling

	Mbd4 vtKO	Mbd4 nKO	Tdg vtKO
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1	62	5
Glutathione Metabolism	2	1	/
VDR/RXR Activation	3	13	90
Ovarian Cancer Signaling	4	/	50
Human Embryonic Stem Cell Pluripotency	5	94	19
Arachidonic Acid Metabolism	6	37	/
Pentose and Glucuronate Interconversions	7	/	168
Hepatic Fibrosis / Hepatic Stellate Cell Activation	8	4	6
LPS/IL-1 Mediated Inhibition of RXR Function	9	3	162
Metabolism of Xenobiotics by Cytochrome P450	10	23	/
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	11	86	2
IGF-1 Signaling	12	52	21
Cysteine Metabolism	13	39	68
Chondroitin Sulfate Biosynthesis	14	46	177
Role of Wnt/GSK-3β Signaling in the Pathogenesis of Influenza	15	/	/
Keratan Sulfate Biosynthesis	16	49	191
Growth Hormone Signaling	17	/	62
Wnt/β-catenin Signaling	18	21	13
Macropinocytosis Signaling	19	/	63
Basal Cell Carcinoma Signaling	20	/	158
Glioblastoma Multiforme Signaling	21	44	3
Axonal Guidance Signaling	22	/	16
Role of JAK2 in Hormone-like Cytokine Signaling	23	/	/
Type I Diabetes Mellitus Signaling	24	16	210
IL-9 Signaling	25	88	/



B. Mbd4 nKO profiling



C. Tdg vtKO profiling

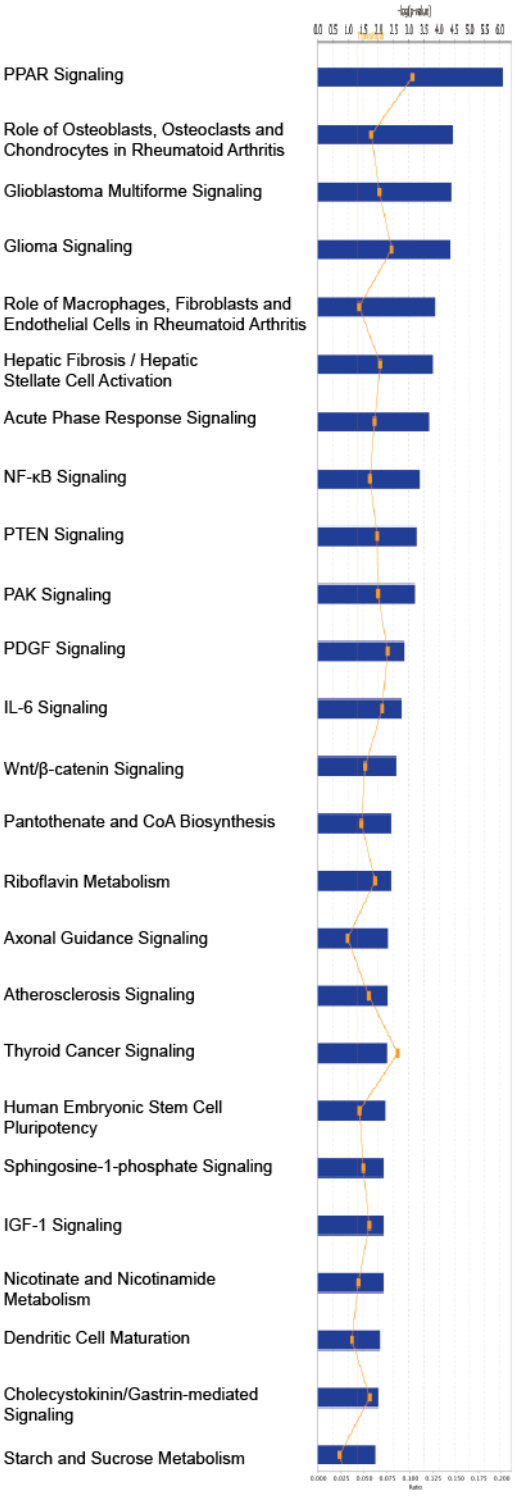
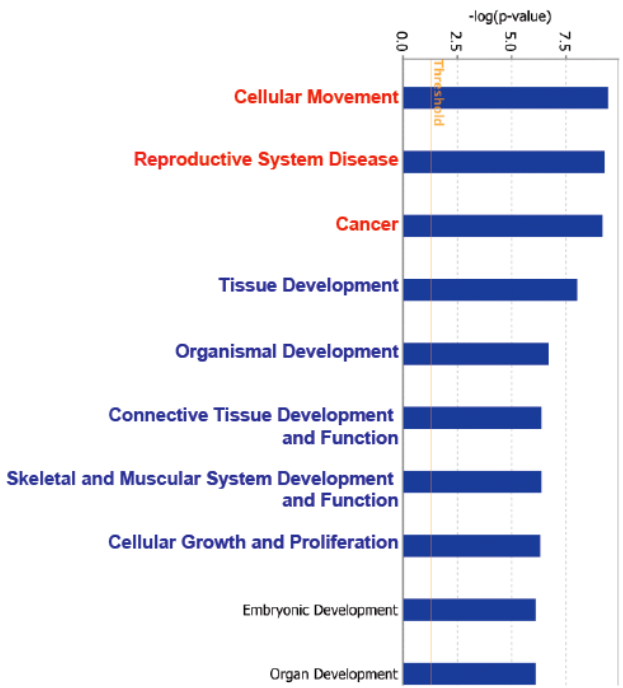


Figure. 5.5. Overlapping canonical pathways identified by analyses of mis-regulated gene expression in the two independent *Mbd4*$^{-/-}$ MEFs (A&B) and transformed *Tdg*$^{-/-}$ MEFs (C). Up- and down- regulated genes with their expression fold changes were functionally analysed by IPA pathway software. The numbers in the lists represent their rankings of indicated canonical pathways in the individual cell lines, based on their p-value significance (The higher –log (p value) number, the more significant). The top 25 categories in each ranking are most significant ones. The colour of canonical pathway names: Red labeled: overlapping between *Mbd4* vtKO and *Tdg* vtKO, which suggests the functional convergence between MBD4 and TDG; Green labeled: overlapping between two independent *Mbd4* KO cell lines, which suggests the conserved and significant function of MBD4. Transformed *Mbd4*$^{-/-}$ MEFs (*Mbd4* vtKO): fold change (FC) at two; Non-transformed *Mbd4*$^{-/-}$ MEFs (*Mbd4* nKO): fold change (FC) at three; Transformed *Tdg*$^{-/-}$ MEFs (*Tdg* vtKO): fold change (FC) at log 1.2. Bar charts represent the p-value significance, in which Fisher's exact test was used to calculate a p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. (The bigger number at the –log(p-value) scale above represents the more significant category group in the cluster, a threshold line was automatically calculated by the software labelled by the yellow vertical line) The yellow dots represent the ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway whose scale is at below.

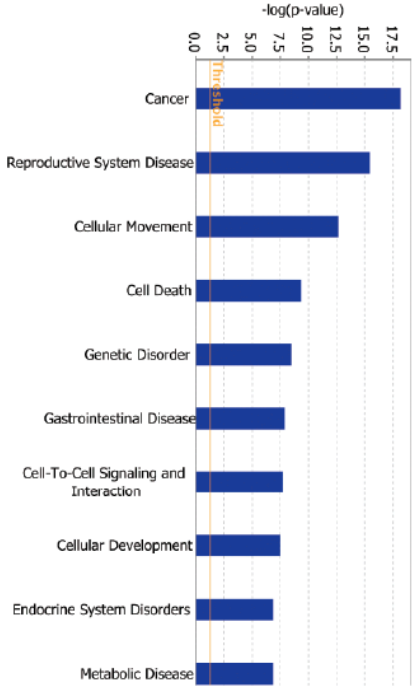
Hepatic Stellate Cell Activation; 4. Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis; 5. IGF-1 Signaling; 6. Wnt/β-catenin Signaling; 7. Glioblastoma Multiforme Signaling; 8. Axonal Guidance Signaling.

The enriched biological functions of MBD4 and TDG were tested by IPA pathway analysis (Figure. 5.6). Seven out of ten significant biological functions of MBD4 and TDG overlapped in all three lists (Figure. 5.6 bold and red or blue colour categories). The top three biological functions identified in both *Mbd4* vtKO and nKO lists were identical (red colour category), which were also in top ten of *Tdg* vtKO list. Except for the top ranking function in cancer, it is very interesting that there are two distinct groups within the other six overlapping biological functions, suggesting additional distinct roles of MBD4 and TDG at different developmental stages. Specifically, the two significant functions of cellular movement and reproductive system disease ranked in top three position in MBD4 functions, which fell to lower rankings, at No. six and eight, in *Tdg* vtKO with less significant p values. In contrast, the less significant biological grouping of development and functions of connective tissue, and skeletal and muscular system in *Mbd4* vtKO were found in top rankings in *Tdg* vtKO, possibly implicating their functionally-compensatable but developmentally-distinct roles in mammals.

Mbd4 vtKO



Mbd4 nKO



Tdg vtKO

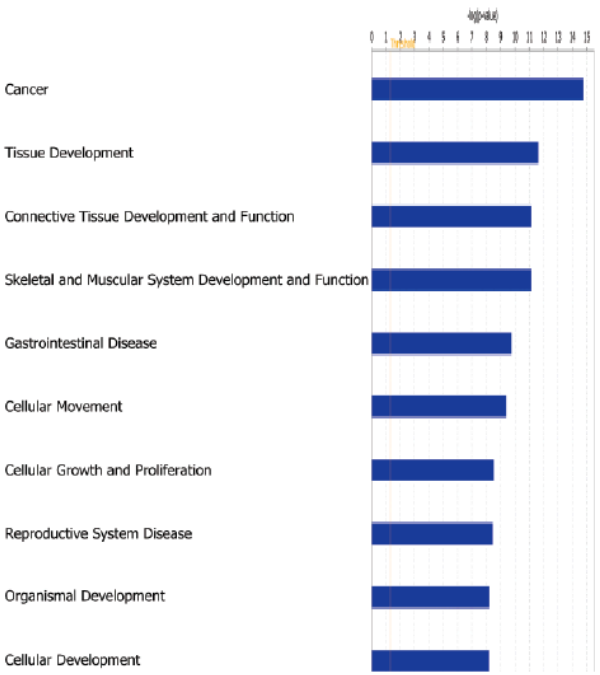


Figure. 5.6. Overlapping biological functions identified by analysis of mis-regulated gene expression in the two independent *Mbd4*$^{-/-}$ MEFs and an additional transformed *Tdg*$^{-/-}$ MEFs. Up- and down- regulated genes with their expression fold changes were functionally analysed by IPA pathway software. The colours of biological function names: Red labeled: overlapping in all three cell lines; Blue labeled: overlapping between *Mbd4* vtKO and *Tdg* vtKO cell lines. Transformed *Mbd4*$^{-/-}$ MEFs (*Mbd4* vtKO): fold change (FC) at two; Non-transformed *Mbd4*$^{-/-}$ MEFs(*Mbd4* nKO): fold change (FC) at three; Transformed *Tdg*$^{-/-}$ MEFs (*Tdg* vtKO): fold change (FC) at log 1.2.

5.2.5. Identification of candidates that are most likely the targets of transcriptional regulation of MBD4 and TDG

The IPA pathway analysis provided a clue of the possible functional significance of MBD4 and TDG. However, the knowledge bases of canonical pathways gathered by IPA have mixed all the known mis-regulated genes of a canonical pathway from a variety of sources of cell types and model systems. For example, the top canonical pathway in *Mbd4* vtKO profiling (Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis) includes three biologically differential cell types. I then asked whether the overlapping gene sets in the 8 overlapping pathway clusters above could be the targets of MBD4 and/or TDG. However, a number of genes including *PDGFB*, *PDGFRA*, *IL33*, *WNT4*, *SFRP1* from the overlapping gene sets in these possible pathway clusters are not consistent in terms of their up- or down- regulation in both cell lines, indicating they are unlikely the convergent targets of MBD4 and TDG regarding their possible function in transcriptional regulation. To identify the specific genes that possibly regulated by MBD4 or TDG (directly or indirectly), I attempted to identify the most significant and consistent gene candidates in all the microarray data of *Mbd4*$^{-/-}$ MEFs and *Tdg*$^{-/-}$ currently available. These include the data from above three cell lines and another independent *Tdg*$^{-/-}$ nKO MEFs cell line (Cortellino et al., 2011). This led to a focus on three significant candidates, which closely correlated with the convergent pathways in Figure. 5.5. These three most interesting candidates are *Col2a1* (collagen, type II, alpha 1), *Crabp2* (cellular retinoic acid binding protein 2), and *Sdpr* (serum deprivation response) (Figure 5.7).

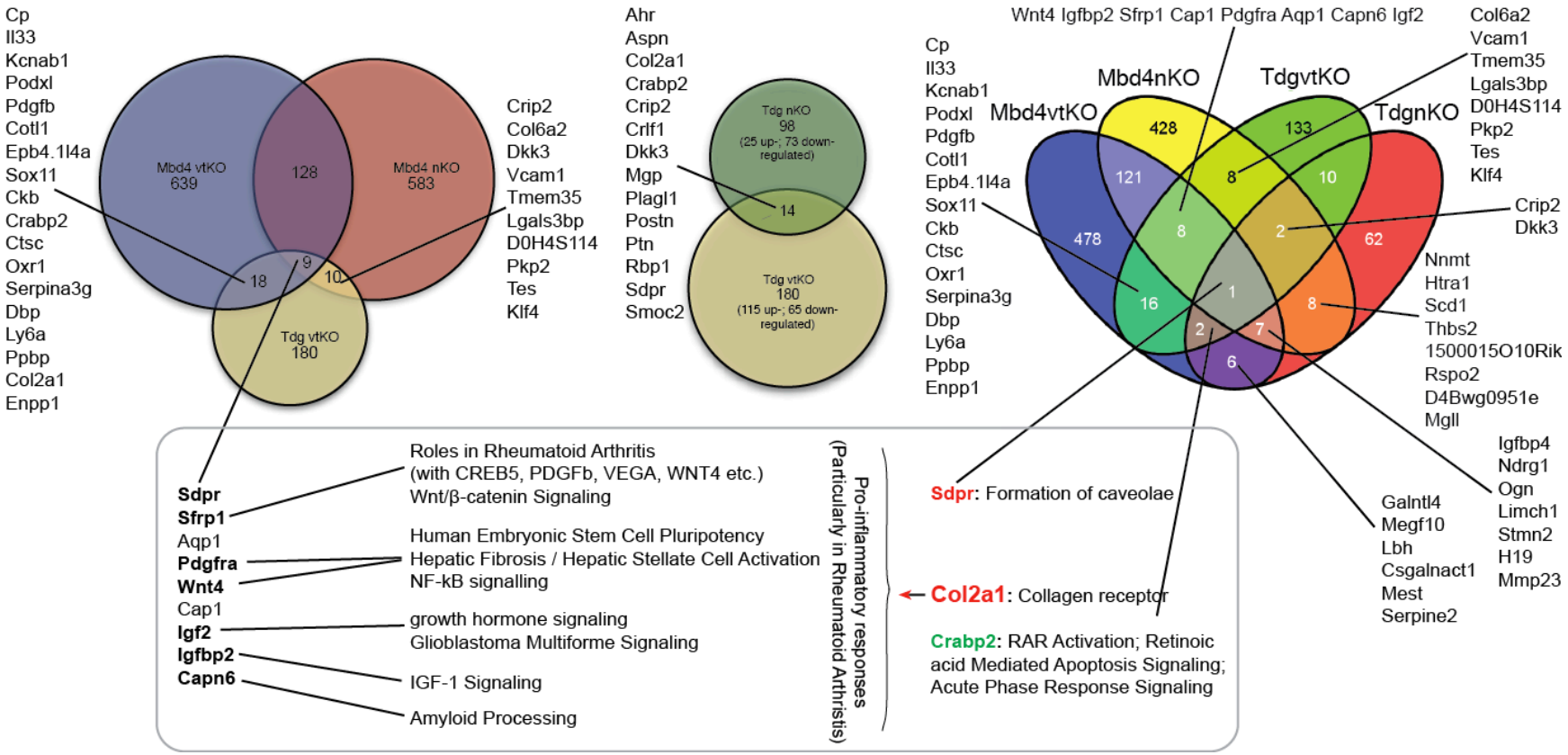
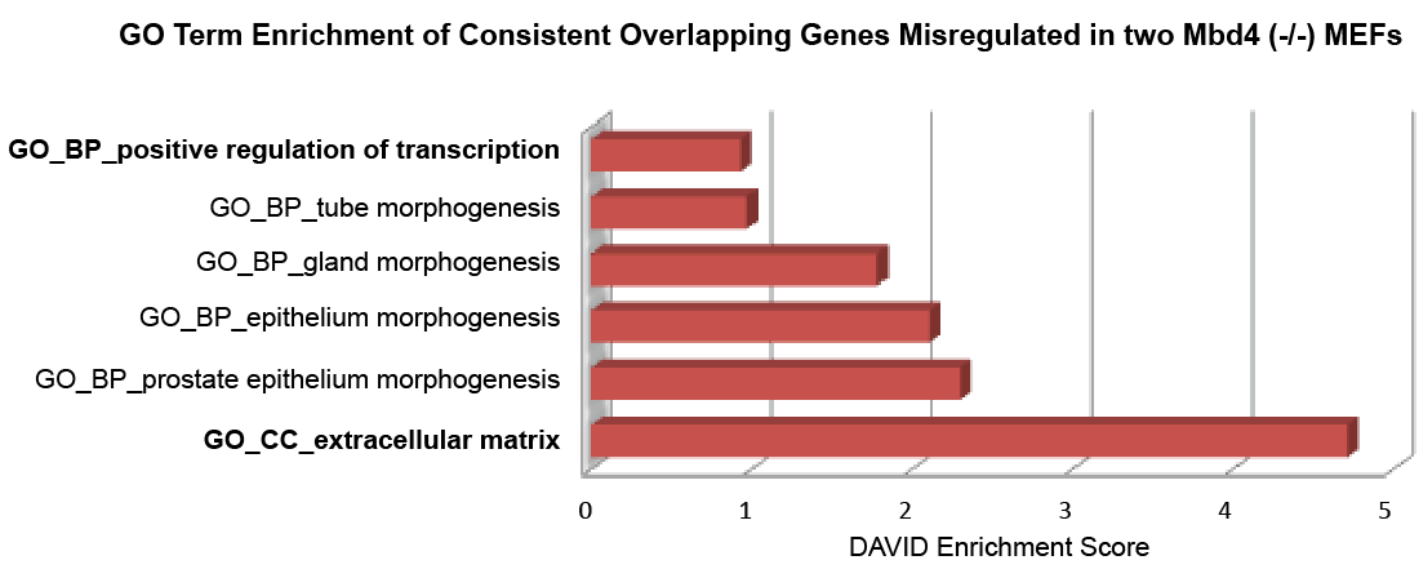

Cp
Il33
Kcnab1
Podxl
Pdgfb
Cotl1
Epb4.1l4a
Sox11
Ckb
Crabp2
Ctsc
Oxr1
Serpina3g
Dbp
Ly6a
Ppbp
Col2a1
Enpp1
Mbd4 vtKO
639
128
Mbd4 nKO
583
18
9
10
Tdg vtKO
180
Crip2
Col6a2
Dkk3
Vcam1
Tmem35
Lgals3bp
D0H4S114
Pkp2
Tes
Klf4
Ahr
Aspn
Col2a1
Crabp2
Crip2
Crlf1
Dkk3
Mgp
Plagl1
Postn
Ptn
Rbp1
Sdpr
Smoc2
Tdg nKO
98
(25 up-; 73 down-regulated)
14
Tdg vtKO
180
(115 up-; 65 down-regulated)
Wnt4 Igfbp2 Sfrp1 Cap1 Pdgfra Aqp1 Capn6 Igf2
Col6a2
Vcam1
Tmem35
Lgals3bp
D0H4S114
Pkp2
Tes
Klf4
Mbd4vtKO
Mbd4nKO
TdgvtKO
TdgnKO
428
133
121
8
10
478
8
2
62
16
1
8
2
7
6
Crip2
Dkk3
Nnmt
Htra1
Scd1
Thbs2
1500015O10Rik
Rspo2
D4Bwg0951e
Mgll
Igfbp4
Ndrg1
Ogn
Limch1
Stmn2
H19
Mmp23
Galntl4
Megf10
Lbh
Csgalnact1
Mest
Serpine2
Sdpr
Sfrp1
Aqp1
Pdgfra
Wnt4
Cap1
Igf2
Igfbp2
Capn6
Roles in Rheumatoid Arthritis
(with CREB5, PDGFb, VEGA, WNT4 etc.)
Wnt/β-catenin Signaling
Human Embryonic Stem Cell Pluripotency
Hepatic Fibrosis / Hepatic Stellate Cell Activation
NF-kB signalling
growth hormone signaling
Glioblastoma Multiforme Signaling
IGF-1 Signaling
Amyloid Processing
Pro-inflammatory responses
(Particularly in Rheumatoid Arthritis)
Sdpr: Formation of caveolae
Col2a1: Collagen receptor
Crabp2: RAR Activation; Retinoic acid Mediated Apoptosis Signaling; Acute Phase Response Signaling

Figure. 5.7. Venn diagram of differentially expressed genes in the two *Mbd4*$^{-/-}$ MEFs and the two *Tdg*$^{-/-}$ MEFs. The names of the cell lines, number of up- and down- regulated genes, and the overlapping gene lists are indicated. The significant overlapping genes implicate for the functional convergence of MBD4 and TDG in the pro-inflammatory responses in cancers and autoimmune diseases such as Rheumatoid Arthritis (collagen induced, *Col2a1*), *Sdpr* (serum deprivation response), and lipid metabolism (RAR activation, *Crabp2*). Red labeled gene: upregulated; Green labeled gene: downregulated.

In addition, to test for enrichment in MBD4 functions Gene Ontology (GO) analysis was undertaken with the 75 consistent overlapping genes between the two independent *Mbd4*$^{-/-}$ cell lines (32 up- and 43 down- regulated) (see method and Figure 5.4 middle). These genes were highly enriched for GO terms involved in cellular components of extracellular matrix, and biological functions in a number of tissue morphogenesis, and transcriptional regulation (Figure. 5.8 A). Most interestingly, in terms of a role of MBD4 in transcriptional regulation, two developmental genes *Sox9* (3.12 fold in vtKO, 3.25 fold in nKO) and *Klf2* (2.65 fold in vtKO, 4.6 fold in nKO) were identified as significantly up-regulated (Figure. 5.8 B), which have been strongly associated with pro-inflammatory responses in a number of acquired conditions such as cancer and autoimmune disease. In contrast, *Sox2*, *Sox11*, and *Klf4* were dramatically upregulated in *Tdg* vtKO MEFs. In addition to this observation, pathway analysis data previously conducted found that *Klf4* might be another target of MBD4 as it was dramatically upregulated (~16 fold) in *Mbd4* nKO cell line (Table 5.2), although no change was observed in another *Mbd4* vtKO MEFs. This was of interest in view of the observations that the non-transformed Mbd4 nKO cell line exhibited a pro-apoptotic phenotype, with cells that show morphological abnormalities, grow more slowly and show senescent-type cell enlargement and spreading (~three or four times normal size) (Figure. 5.1 C), compared with their responsive wildtypes. IPA pathway analysis and Gene Ontology (GO) analysis with the nKO genes only was conducted, and identified a cluster of significant genes implicated in regulation of apoptosis. This included the three developmental genes referred to above, as well as a mixture of apoptosis suppression and induction genes (Table 5.2). Among these genes, *Klf4* (~16 fold upregulated in nKO) and *Cdkn1a* (~17 fold upregulated in nKO, 1.5 fold in vtKO) are most significant, suggesting they might play an important role in suppressing apoptosis and inducing cell-cycle arrest in the cells lacking MBD4 activity.

A.



B.

GO_BP_positive regulation of transcription:	SOX9, KLF2, SDPR	BCL3, HOXB6, KCNH2, PPARGC1B
GO_CC_extracellular matrix:	COL8A1, COL2A1, TNC, PRELP	FBLN1, WNT4, WNT5B, SMOC1, SPON2
	Upregulated genes	Downregulated genes

Figure. 5.8. GO enrichment analysis (see Methods) of genes that misregulated in both two *Mbd4*$^{-/-}$ MEFs. (A) The overlapping genes of *Mbd4* vtKO (fold change 2) and *Mbd4* nKO (fold change 3) with consistent up- or down- regulated changes in both cell lines were imported into DAVID GO term analysis. The top clusters are listed, including a top ranked cellular component of extracellular matrix cluster, and another 5 biological process clusters. (B) The genes in two clusters of transcriptional regulation and extracellular matrix are listed, with their up- or down- regulation status. The other process clusters in (A) contain different combinations of the genes listed in (B).

Gene ID	vtKO	nKO	Gene Name
Cdkn1a	1.5	17.26	cyclin-dependent kinase inhibitor 1A (P21)
Klf4	1	15.77	Kruppel-like factor 4 (gut)
Adrb2	-2.94	11.74	adrenergic receptor, beta 2
Cryab	-5.88	9.07	crystallin, alpha B
Nod1	-3.85	5.04	nucleotide-binding oligomerization domain containing 1
Nupr1	1.12	4.8	nuclear protein 1
Klf2	2.65	4.6	Kruppel-like factor 2 (lung)
Bdnf	x	4.23	brain derived neurotrophic factor
Ngfrap1	1.2	4.14	nerve growth factor receptor (TNFRSF16) associated protein 1
Trp53inp1	1.44	4.12	transformation related protein 53 inducible nuclear protein 1
Phlda3	-1.37	3.88	pleckstrin homology-like domain, family A, member 3
Gsk3b	1.26	3.79	glycogen synthase kinase 3 beta
Fas	1.08	3.74	Fas (TNF receptor superfamily member 6)
Bax	1.24	3.34	BCL2-associated X protein
MDM2	1.16	3.31	transformed mouse 3T3 cell double minute 2
Hes1	-1.15	3.28	hairy and enhancer of split 1 (Drosophila)
Sox9	3.12	3.25	SRY-box containing gene 9
Tle1	-1.75	2.96	transducin-like enhancer of split 1, homolog of Drosophila E(spl)
Kras	1.4	2.85	similar to GTPase KRas precursor (K-Ras 2) (Ki-Ras) (c-K-ras) (c-Ki-ras); v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
Casp7	-1.09	2.61	caspase 7
Dcun1d3	1.13	2.32	DCN1, defective in cullin neddylation 1, domain containing 3 (S. cerevisiae)
Tnfaip3	-2.17	2.26	tumor necrosis factor, alpha-induced protein 3
Mapk8ip1	1.57	-3.8	mitogen-activated protein kinase 8 interacting protein 1
Tle2	1.18	-4.16	transducin-like enhancer of split 2, homolog of Drosophila E(spl)
Bcl3	-2	-4.55	B-cell leukemia/lymphoma 3
Bcl2l11	-1.2	-4.76	BCL2-like 11 (apoptosis facilitator)
Stk3	1.21	-5	serine/threonine kinase 3 (Ste20, yeast homolog)
Gch1	-1.15	-5.26	GTP cyclohydrolase 1
Cln3	8.62	-5.88	similar to CLN3; ceroid lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)

Table. 5.2. Cluster of significant genes in terms of regulation of apoptosis in nKO cell line. Fold changes of up- or down- regulated genes in *Mbd4* vtKO and nKO are labelled. Red labelled: significant upregulated genes in nKO cell line; Green labelled: significant downregulated genes in nKO cell line. Yellow marked: top significant genes.

5.3. Discussion

The purpose of transcriptional profiling of independent *Mbd4*$^{-/-}$ MEFs cell lines was to identify the possible biological relevance of MBD4 in transcriptional regulation. Illumina mouse WG-6 bead arrays identified a large group of genes that directly or indirectly regulated by MBD4 activity. Additionally, I included the currently available data of the only other mammalian thymine glycosylase *Tdg*$^{-/-}$ MEFs from the same platform of Illumina beadarray. By functional annotation and pathway analysis, I showed that MBD4 might have functions in the regulation of the hormone induced nuclear receptors. In addition, MBD4 and TDG might have convergent functions in a number of significant canonical pathways, possibly in collagen induced pro-inflammatory responses in Rheumatoid Arthritis (RA) and cancer. Finally, several developmental genes including *Sox9*, *Klf2*, and *Klf4*, and paternally imprinted genes *Dio3* and *H19*, were identified as possible targets of transcriptional repression by MBD4.

5.3.1. A possible direct interaction between MBD4 and ERα

The interaction of MBD4 and ERα suggests that MBD4 might potentiate transcription of Estrogen regulated genes via a direct interaction with ERα, supporting a possible link between replication repair remodelling and steroid hormone receptor transcriptional regulation. Interestingly, the other mammalian G:T mismatch-specific thymine-DNA glycosylase TDG was also reported to potentiate transcription of Estrogen-regulated genes through direct interaction with Estrogen receptor α (Chen et al., 2003). The protein sequences responsible for the interaction with ERα lie within the glycosylase domain of TDG (Chen et al., 2003), suggesting a possible similar way that the C-terminal glycosylase domain of MBD4 might interact with ERα. However, a mutant TDG-N140A, which lacks glycosylase activity, can still stimulate ERα activity at the same extent of wild-type TDG, suggesting the DNA glycosylase activity is not required for stimulation of ERα activity by TDG (Chen et al., 2003). Instead, TDG may recruit co-activators such as CBP/p300 and the

p160 to stimulate the transcriptional activity, and to facilitate base excision repair by histone acetylation (Tini et al., 2002, Rai et al., 2010). In Chapter 3 &4, I identified and validated a large number of transcriptional co-regulators that interact with MBD4, many of them associated with MBD4 regardless of its glycosylase domain. Thus it is very possible that MBD4 acts in a similar way to TDG, regulating gene expression in a ligand-dependent manner in the context of steroid hormone receptors such as ERα. However, whether and how these steroid hormone receptors regulate DNA repair by MBD4 and/or TDG, possibly facilitated by the recruitment of chromatin-remodeling complexes, and/or in a context of DNA methylation marks including 5mC and the newly discovered 5hmC, still remain to be elucidated.

Interestingly, in a number of top functional implications of canonical pathway annotations, MBD4 was implicated in several adopted orphan nuclear receptors (RXR)/the RXR heterodimer receptors pathways including VDR/RXR Activation, LPS/IL-1 Mediated Inhibition of RXR Function; and Metabolism of Xenobiotics by Cytochrome P450 in both two independent *Mbd4*$^{-/-}$ cell lines (Figure. 5.5 A). This suggests that MBD4 might act as a possible co-regulator at these non-steroid nuclear receptors, to facilitate the lipid homeostasis. Consistent with this possible significance of MBD4 from these annotations, Kim et al. demonstrated a phenomenon of active DNA demethylation at a specific promoter of the cytochrome p450 27B1 (CYP27B1) gene in vitamin-D hormone induced transcriptional derepression/activation (Kim et al., 2009b). It is likely that in these lipid metabolic processes, MBD4 might firstly directly target these adopted orphan nuclear receptors (RXR)/the RXR heterodimer receptors, and subsequently activate them by possible active demethylation routes of removal of deamination products such as thymine or the oxidative products such as hydroxymethylcytosine by its glycosylase repair function when induced by the responsive hormones. TDG might act in a similar way regarding these non-steroid hormone receptors. The PPAR signalling was identified as the top canonical pathway in annotation of *Tdg* vtKO (Figure. 5.5 C), which concurs with previous reports that both the adopted orphan nuclear receptors (retinoid receptors including all-trans-retinoic acid α and 9-cis-retinoic acid α), and the other heterodimers receptors (PPARγ, thyroid hormone α, and vitamin D3) interact with TDG in a ligand-independent manner (Um et al., 1998, Chen et al., 2003).

5.3.2. Possible functional convergence between MBD4 and TDG in acquired conditions such as Rheumatoid Arthritis (RA), and cancer.

Unexpectedly, pathway analysis implied a possible functional convergence of MBD4 and TDG in pro-inflammatory responses, especially in autoimmune disease such as Rheumatoid Arthritis (RA). In addition, it is interesting that the other overlapping pathways in this top list correlate with the long-term complications that clinically observed in patients with Rheumatoid Arthritis. For example, one of the systemic consequences of RA is the increased rates of cardiovascular illness, including myocardial infarction, cerebrovascular events, and heart failure, which are not explained by traditional risk factors, use of glucocorticoids, nonsteroidal anti-inflammatory drugs, or shared genetic features (McInnes and Schett, 2011). The significant pathway implied in Figure 5.5 such as Wnt/β-catenin Signaling, and a number of mis-regulated transcripts such as *SOCS3*, *TNFSF11*, *PDGFB*, *VEGFA*, are true cardiovascular risk factors. In addition, other systemic complications of RA including liver (elevated acute-phase response and anaemia of chronic disease), lungs (inflammatory and fibrotic disease, muscle (sarcopenia), and brain (fatigue and reduced cognitive function) could be associated with the identified overlapping pathways of hepatic fibrosis / hepatic stellate cell activation, human embryonic stem cell pluripotency, IGF-1 signaling, glioblastoma multiforme signaling, and axonal guidance signaling.

Rheumatoid Arthritis (RA) is a common systemic inflammatory disorder that primarily affects the small diarthrodial joints of the hands and feet, and is associated with progressive disability, systemic complications, and early death. (Firestein, 2003, McInnes and Schett, 2011) The clinical features of RA include synovial inflammation and hyperplasia, autoantibody production, cartilage and bone destruction, and systemic features, including cardiovascular, pulmonary, psychological, and skeletal disorders. (McInnes and Schett, 2011) The underlying cause of RA is unknown, although recent studies have associated RA with a number of genetic and environmental factors that influence T-cell activation and NF-κB pathway. RA shows a complex interplay among these genetic and environmental factors, the drugs and therapies targeting these pathways sometimes fail or produce only partial responses (McInnes and Schett, 2011). The possible functional convergence of MBD4 and TDG in role of RA suggests they might be candidates as predictive biomarkers of diagnosis and prognosis of RA. The current classification of RA is based on the clinical phenotype (Aletaha et al., 2010). A variety of RA related canonical pathways have been implicated in the functional annotations of *Mbd4*$^{-/-}$ and *Tdg*$^{-/-}$ microarray results, suggesting it is possible that MBD4 and/or TDG might be the candidates for new

molecular taxonomy markers that represent subgroups within RA: this requires further investigation.

In a widely used model of rheumatoid arthritis (RA), murine collagen induced arthritis (CIA), MBD4 was identified as one of the five top susceptibility genes controlling arthritis development (Yu et al., 2009). This collagen-induced arthritis (CIA) mouse model is the most commonly studied autoimmune model of rheumatoid arthritis and is thought to reflect many features of human disease pathogenesis. Autoimmune arthritis is induced in this model by immunization with an emulsion of complete Freund's adjuvant and type II collagen (CII) (Brand et al., 2007). Yu et al. investigated six genomic regions that showed suggestive level of linkage to CIA in F2 mice. They used the advanced intercross line (AIL) mice to confirm those suggestive quantitative trait loci (QTL) and refine the positive(s) into a small genomic region, thus identifying MBD4 as a top candidate with the nonsynonymous polymorphisms and differentially expressed (Yu et al., 2009). Moreover, two polymorphism mutations of MBD4 were statistically associated with patients with RA in Chinese population (Huang et al., 2010a). It is noteworthy that the key transcripts that impact the collagen induced arthritis pathways were mis-regulated in the *Mbd4*$^{-/-}$ and *Tdg*$^{-/-}$ cell lines, including *Sfrp1*, *Pdgfra*, the Type II collagen receptor *Col2a1*, and a cohort of NFκB activating transcripts (Figure. 5.7 & supplementary Figure. 5.1 B). Therefore, it is interesting to validate this possible biological significance of MBD4 in RA using murine collagen induced arthritis (CIA) model, by treating *Mbd4*$^{-/-}$ mice with the immunization with an emulsion of complete Freund's adjuvant and type II collagen (CII), to confirm if the depletion of *Mbd4* could lower the threshold of the induction of RA. In addition, it is also very interesting to investigate several crucial factors in RA that were coincidently identified for their significance in the network analyses including NF-κB pathway associated factors, apoptosis resistance factors, and the interaction with stress proteins (e.g., heat-shock protein 70). In addition, the apoptosis role of MBD4 needs to be further investigated in the context of RA, as a likely possibility that sustain synovial hyperplasia would be the altered resistance to apoptosis. Thus it will be very interesting to further study the apoptosis role of MBD4, regarding the link between the P53 and the NF-κB pathways.

The biological functions of MBD4 and TDG annotated by IPA analysis reinforced the their possibility that their functions converged (Figure. 5.6). Interestingly, their implied biological role in cancer was the top association. This is perhaps unsurprising because the possible pro-inflammatory responses caused by lack of MBD4 or TDG activity might predispose the affected organs towards cancer for example in colon and lung. In addition, MBD4 and TDG are both thymine glycosylases that can excise DNA mismatches derived

from deamination or oxidative stress; loss of such repair activity increase the the risk of tumorigenesis when combined with other oncogenic stimuli (Millar et al., 2002). The other two overlapping biological functions identified were cellular movement and reproductive system disease (Figure. 5.6), suggesting MBD4 and TDG may have significant roles in these functions. Notably, although several key biological function groups were top ranked in both MBD4 and TDG functions, their priorities in these two glycosylase proteins were quite distinct. Three key developmental functions including Tissue Development, Connective Tissue Development and Function, and Skeletal and Muscular System Development and Function were ranked as very top functions in TDG list (Top No. 2, 3, 4), whereas they were demonstrated less significant in MBD4 list (Figure. 5.6). In contrast, the two top functions in MBD4 list (cellular movement and reproductive system disease) are at a much less significant position in the TDG list. The highly complementary but distinct prioritised biological functions between MBD4 and TDG might explain in part regarding the distinct phenotypes of knockout mice that lack MBD4 or TDG. Although MBD4 and TDG might compensate each other in some regards, the developmental role of TDG may be much more significant than that of MBD4, thus the mice lacking TDG were lethal at early developmental stage, whereas mice with *Mbd4* deficiency were viable. Presumably when induced by stress factors such as arthritis inducing collagen, oxidative damage, other oncogenic stimuli, or aging, the *Mbd4*$^{-/-}$ mice would show a much lower threshold to develop overt disease.

5.3.3. Sox9, Klf2, and Klf4: possible targets of MBD4 regulation in pro-inflammatory responses and pro-apoptotic responses in human breast cancer

One of the most interesting aspect of my gene expression analyses was the identification of several developmental genes including *Sox9*, *Klf2*, and *Klf4* as possible targets of MBD4, in terms of its possible roles in pro-inflammatory (*Sox9*, *Klf2*, *Klf4*) and pro-apoptotic responses (*Klf4* and its signalling through *Cdkn1a* (the gene that encodes p21)) in acquired conditions such as certain types of cancer and immune disease. Interestingly these possible targets of MBD4 have been reported to primarily expressed in certain tissues, for example *Sox9* in testis, pancreas, intestine, brain, kidney, heart valves and derivatives of the neural crest (Pritchett et al., 2011), *Klf2* in lung, blood vessels, lymphocytes, endothelial cells (Suzuki, 2005), and *Klf4* in vascular smooth muscle cells, endothelial cells and gut-enriched (Suzuki, 2005). The expression of a vast number of target genes are regulated by SOX9,

KLF2 and KLF4, thus it is not surprising that they are involved in many cellular functions, ranging from differentiation to proliferation and apoptosis (Rowland and Peeper, 2005).

SOX9 (Sex-determining Region Y (SRY) box 9) play crucial and diverse roles during development (Pritchett et al., 2011). It belongs to a group of transcription factors that sharing the high mobility group DNA-binding domain of SRY (Gordon et al., 2009). In addition to its developmental role at many sites including chondrogenesis and testis formation, SOX9 is also expressed and implicated in the formation and growth of cancers derived from several different cell types, including prostate, brain, colon, and skin (Pritchett et al., 2011). The roles of SOX9 in chondrogenesis and fibrosis, sclerosis and related disorders have been well associated with the mis-regulation of a number of extracellular matrix (ECM) (Pritchett et al., 2011, Leung et al., 2011), many of which coincide with the ECM genes identified from my GO term analysis (Figure. 5.8), including the common target *Col2a1* (Figure. 5.7&5.8). In this context, it is very possible that Sox9 might be a key target of MBD4 regulation in terms of its role in autoimmune diseases such as RA implicated from the above pathway analysis. In addition, it is interesting that SOX9 seems to be important for epithelial-to-mesenchyme transition (EMT) with EMT-related transcription factor Snail2 (Slug) in chick neural tube (Cheung et al., 2005). Most interestingly, a recent report has found that SOX9 and Slug suffice to convert differentiated epithelial cells into long-term repopulating epithelial stem cells in a human mammary cell line (Guo et al., 2012), indicating that SOX9 is a key transcriptional regulator in such EMT transition. This might underline the associations between the high expression level of SOX9 and the poor prognosis of a range of tumors including colon cancer, prostate cancer (Pritchett et al., 2011), and especially human breast cancer (Guo et al., 2012).

KLF2 and KLF4 are Kruppel-like factors, a subclass of the zinc finger family of transcription factors characterised by the three Kruppel-like zinc fingers (Rowland and Peeper, 2005). The KLF proteins bind to common GC-rich DNA sequences, suggesting that their specificity depends on their co-operators and in terms of the certain tissues they express, and different KLFs might compensate each other to coordinately regulate the expression of KLF effector genes (Rowland and Peeper, 2005). Indeed, It seems the mild endothelial phenotype of *Klf2*$^{+/-}$ mice was due to the compensatory upregulation of *Klf4* (Atkins et al., 2008), and KLF4 strongly resembles KLF2 at least in the vascular endothelium (Kunes et al., 2009). Both KLF2 and KLF4 were reported to have activity in the vascular endothelium, and are induced by laminar shear stress that provokes the expression of various protective factors (Kunes et al., 2009). Interestingly these are counter-forces of pro-inflammatory factors induced by inflammatory stimuli such as NF-κB. In the context of

infection or wound healing, the pro-inflammatory responses and protective responses need to be fine tuned, as although the activation of inflammation helps to combat infection and renew organ integrity, the prolonged activation out of control may degrade into autoimmune diseases or cancer growth (Kunes et al., 2009). In addition to the induction of protective factors, KLF4 is required for induction of p21 (CDKN1A) in a p53 dependent manner in response to DNA damage, which results in cell-cycle arrest (Rowland and Peeper, 2005). This is particularly interesting considering that KLF4 overexpression was found in up to 70% of primary human breast cancers (Rowland and Peeper, 2005), suggesting that there might be a oncogenic role of KLF4 through p21-dependent cell-cycle arrest at least in this genetic context.

MBD4 itself is a sensor for induction of apoptosis possibly through its association with DNMT1 (Ruzov et al., 2009), and/or Fas pathway protein FADD (Screaton et al., 2003). Thus it is not surprising that the removal of MBD4 might result in disruption of normal apoptosis, a phenomenon that is key to the imbalance of death and proliferation that is the hallmark of cancer and immune homeostasis. Consistent with this idea, highly upregulated *Klf4* and *CDKN1A* were identified in one of the *Mbd4* knockout cell line, which could be related to the proposed roles for the inactivation of *Mbd4* in suppressing apoptosis and inducing cell-cycle arrest. In the certain setting that pro-inflammatory responses are out of control caused by the unbalance of pro-inflammatory (and/or proliferative) factors (regulated by SOX9) and protective factors (regulated by KLF2 and KLF4), the misregulation /or disruption of normal apoptosis (regulated by KLF4 and CDKN1A) might prolong such abnormal status. Conversely such unbalanced status of normal tissue might become an 'optimal balance' between continued proliferation and sensitization to apoptosis, which is beneficial for effective tumour expansion and progression of autoimmune diseases. These would particularly match the theme in mammary diseases such as human breast cancers, where the normal terminal duct lobular unit (TDLU) is a dynamic structure that undergoes cyclical changes including epithelial proliferation and apoptosis during the menstrual cycle (Rubin and Strayer, 2012), thus conceivably it is not surprising that unbalance of these processes could result in a 'disease status'. Moreover, it is very interesting that the nuclear localisation of KLF4 has been associated with the poor outcome of breast cancer (Pandya, 2004), and overexpression of Klf4 is associated with the stage of ductal carcinoma in situ (Foster et al., 2000), an early event in breast cancer progression (Rowland and Peeper, 2005). This raises an immediate question that if MBD4 possesses a signalling role, which could switch *Klf4* to an oncogenic status (possibly depending on the status of p21 (Rowland and Peeper, 2005)) , in certain settings such as human breast cancers.

It will be very interesting to know the co-incidence of nuclear localisation of MBD4 and KLF4, which might underline such associations of poor outcome of breast cancer.

5.3.4. Imprinting control regions of H19 and Dio3: possible targets that MBD4 regulates?

In higher mammals, imprinted genes are transcribed from only one parental allele, and play key developmental roles and are generally controlled by imprinting control regions (ICRs) containing CpG-rich sequences (Quenneville et al., 2011). ICRs can affect the expression of imprinted genes organised in clusters through various mechanisms (Quenneville et al., 2011). Of these mechanisms, DNA methylation is the key epigenetic modification known to mark the two parental chromosomes at the differentially methylated regions (DMRs) (Ferguson-Smith, 2011). This results in an epigenetic paradigm of cis-acting mechanisms of gene regulation, which can act over short or long distances, and associate the epigenetic state, the chromatin structure to genome function (Ferguson-Smith, 2011). In particular, two imprinting gene clusters containing H19 and Dio3 respectively are paternally methylated (Ferguson-Smith, 2011) (Figure.5.9). Their DMRs are both intergenic and paternally methylated. This intergenic DMR is needed for normal imprinting of H19, and the DMR that partially overlaps with H19 promoter becomes methylated, associating with the repression of H19 gene (Figure. 5.9. H19). In contrast, the intergenic DMR at Dlk1-Dio3 domain is far from the Dio3 gene, and the methylated DMR at paternally inherited chromosome is generally associated with the expression of Dio3 by an unknown mechanism (Rocha et al., 2008) (Figure. 5.9. Dio3). It is very interesting that in my microarrays profiling two independent *Mbd4*$^{-/-}$ MEFs, H19 was significantly upregulated (eight fold in vtKO, three fold in nKO), whereas Dio3 was dramatically downregulated (seven fold in vtKO, forteen fold in nKO) in both cell lines (All maternally imprinted genes either have no change or are undetectable in the assays). Considering the methyl-CpG binding ability of MBD4, and I have identified a cohort of its association co-repressive complexes, it is very likely the silence of H19 and expression of Dio3 in wildtype MEFs may depend on the MBD4 association co-repressive complexes at their DMRs (Figure. 5.9. A).

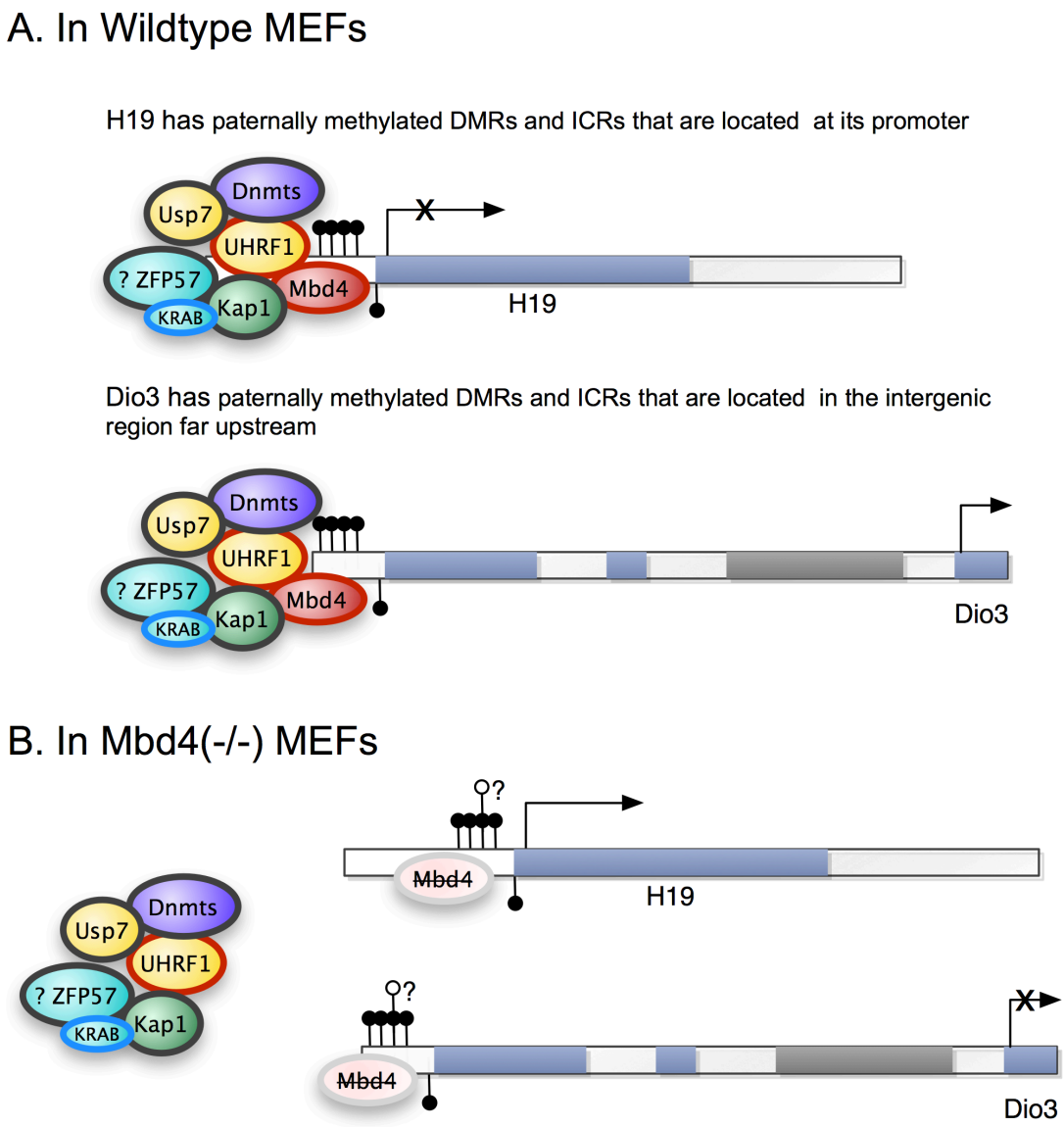


Figure. 5.9. A working model of possible role of MBD4 in regulation of paternally imprinted genes. As indicated, *H19* has paternally methylated DMRs and ICRs that are located at its promoter. In contrast, Dio3 has paternally methylated DMRs and ICRs that are located in the intergenic region at far upstream. (A) The silence of *H19* and expression of *Dio3* in wildtype MEFs may depend on the MBD4 association co-repressive complexes at their DMRs. (B) The deassociation of these repressive complexes may happen in MEF cells that lacking MBD4 activity, which may result in the high expression of *H19*, and dramatic inactivation of *Dio3*. The methylation status at their DMRs might largely keep unchanged, but need further investigation.

In terms of the MBD4 association co-repressive complexes, the association of ZFP57/KAP1 complex and DNMTs and UHRF1 has been recently reported to recognize a methylated hexanucleotide of TGCCGC motif to affect chromatin and DNA methylation of

imprinting control regions in a parental allele-specific manner, and protect the specific loci against DNA demethylation in embryonic stem cells (ESCs) (Quenneville et al., 2011). There are several evidences in line with this mechanism. Firstly, *ZFP57* gene knockout experiments in mice and loss-of-function mutations in human individuals affected by transient neonatal diabetes have recently implicated the role of ZFP57 in the establishment and maintenance of several imprinted loci (Mackay et al., 2008, Li et al., 2008). Secondly, *Kap1* inactivation results in a loss of heterochromatin marks at ICRs, and in the mutants KRAB-ZFPs no longer mediate transcriptional repression (Li et al., 2008, Mackay et al., 2008). Thirdly, the Kruppel-associated box repressor domain recruits KAP1 to the vicinity of promoters and induces their promoter methylation during mouse early embryogenesis (Wiznerowicz et al., 2007), whereas deletion of ZFP57 or DNMTs results in ICR DNA demethylation in ESCs (Quenneville et al., 2011). Last but not least, *Kap1* deletion in ESCs results in markedly increased H3K9ac and depletion of H3K9me3 at ICRs of many imprinted genes, coinciding with the misregulation of similar set of imprinted genes and depletion of *Kap1* and H3K9me3 in *Zfp57*$^{-/-}$ ES cells (Quenneville et al., 2011).

The expression of imprinted genes, which is parental allele-specific, has been remarkably associated with asymmetric chromatin and DNA methylation signatures at ICRs (Quenneville et al., 2011). After the stage of pluripotency, these loci need to be well maintained in a repressed status, where the proper asymmetric histone modifications, heterochromatinization, and DNA methylation at these elements shall be kept by a sequence-specific manner by repressive protein complexes. The evidence that the ZFP57-dependent recruitment of KAP1 at ICRs is required for the maintenance of epigenetic signatures at DMRs of imprinted genes in ESCs, coincides with my protein interaction findings that the same set of proteins including KAP1, UHRF1, and DNMTs interact directly with MBD4, and two of only four paternally imprinted genes H19 and Dio3 were dramatically mis-regulated in *Mbd4*$^{-/-}$ MEFs (the other 2 paternally imprinted genes were undetectable in my microarrays). In addition, interestingly another MBD4 interacting protein PRMT5 has been reported to interacts with and depends on the corepressor KAP1 and KRAB-A domain of ZNF224 to perform the chromatic modification at the human aldolase A gene (Lupo et al., 2011). It seems there might be different combinations between the same repressive complex containing KAP1 and its associated proteins and several KRAB-ZFPs to maintain the transcriptional repression at specific loci in different cellular contexts. An interesting report was that in mice depleted for *Kap1* in the adult forebrain, deregulation of imprinted genes were found in the transcripts of the hippocampus, where ZFP57 is expressed (Jakobsson et al., 2008), suggesting a possible role of ZFP57 and KAP1 complex in controlling imprinting

outside of the early embryonic and germ cells. However, if this is a direct effect and if this really happens in adult somatic cells need further investigations. In sum, it is very likely that MBD4 may have a role in the maintenance of the repression status at DMRs of the paternally imprinted genes such as H19 and Dio3, associating with its co-repressive complexes (Figure. 5.9).

5.3.5. Possible role of MBD4 in human cancers

The inactivation of *MBD4*, either by mutations (Bader et al, 1999, Riccio et al, 1999, Menoyo et al, 2001, Yamada et al, 2002, Bader et al., 2007) or hypermethylation at its promoter region (Howard et al., 2009) have been reported and proposed to associate with certain types of human cancers including colorectal cancers, ovarian cancers and liver cancers (Saito et al., 2001). All of these types of *MBD4* inactivation event occur in parallel with the tumors exhibiting microsatellite instability (MSI), caused by inactivation of mismatch repair (MMR) protein. However, whether the *MBD4* inactivation could affect early tumorigenesis, tumour progression, and/or metastasis, are largely unknown.

In my study, I have identified a cohort of MBD4 associated proteins, including the three co-repressor proteins KAP1, PRMT5 and UHRF1, which interact with MBD4 at its intervening domain. Interestingly naturally occurring MBD4 mutations are usually a frameshift in two polyadenine tracts resulting in truncated MBD4 proteins at codons 310–313 and 247–248, respectively (Bader et al., 2007, Lucci-Cordisco and Neri, 2009), which would reduce the ability of MBD4 to interact with the above co-repressors. This may be well linked to the dramatic upregulation of the possible MBD4 targets *Sox9*, *Klf2*, and *Klf4* in *Mbd4*$^{-/-}$ MEFs, in which the inactivation of *Mbd4* might result in the derepression of these specific genes, by dissociation of MBD4 interacting co-repressors possibly at or near their promoter regions. Interestingly the proximal 5' flanking region of *SOX9* contains binding sites for the NF-κB complex subunit RelA (Ushita et al., 2009) and CREB consensus motif Sp1 site (Piera-Velazquez et al., 2007), which are likely to be the potential binding sites of MBD4. The coupling of SOX9 and coactivators of CREB binding proteins CREB/p300 can transcriptionally regulate a number of specific genes including the extracellular matrix genes (Pritchett et al., 2011). Consistently, a common feature of some KLFs including KLF2 and KLF4 is that they share the same CBP/p300 (Suzuki, 2005). It is interesting that their close relative protein KLF5 has also been reported to be co-activated by both CBP/p300 (Miyamoto et al., 2003, Zhang and Teng, 2003) and p50 subunit of NF-κB complex (Aizawa et al., 2004), suggesting a common mechanism of transcriptional regulation within the KLFs

through these activation complexes. An oncogenic regulator SET and a histone deacetylase HDAC1 have been shown to negatively regulate transcriptional activity of KLF5 through direct interaction as well as inhibition of its interaction with p300 (Matsumura et al., 2005, Miyamoto et al., 2003). These observations suggest the positive activation and negative regulation of KLFs (possibly Sox9 as well) and their downstream genes may depend on the different cofactors coupled with them in different conditions. A rational explanation of the coincidence of the robust upregulation of *Sox9*, *Klf2*, and *Klf4* in the MEF cells that lacking MBD4 activity would be due to the disassociation of MBD4 coupling co-repressors (Figure.5.10. I & II). The zinc finger domain of zinc finger transcription factors can form homodimers or heterodimers through protein-protein interactions with themselves as well as other cofactors (Suzuki, 2005, Mackay and Crossley, 1998) including two MBD4 interaction proteins KAP1 and PRMT5, and the different recruitment of co-factors may largely modulate these genes in different tissues and developmental contexts. Thus it is very likely the different zinc finger proteins might form heterodimers with MBD4 and its associated protein complexes in different contexts, to help the recruitment of the complexes to the targeting promoter binding sites.

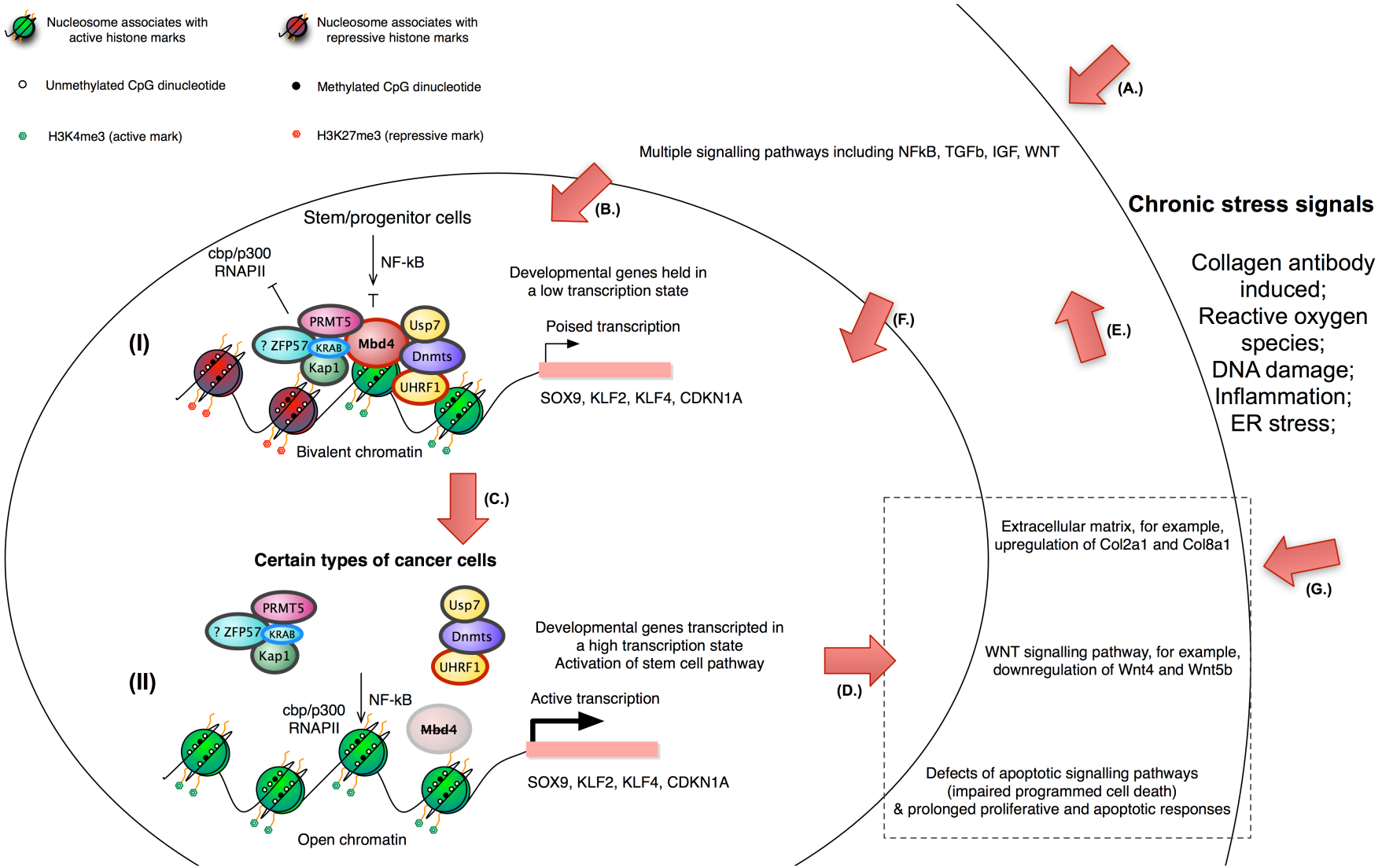

Nucleosome associates with active histone marks
Nucleosome associates with repressive histone marks
Unmethylated CpG dinucleotide
Methylated CpG dinucleotide
H3K4me3 (active mark)
H3K27me3 (repressive mark)
(A.)
Multiple signalling pathways including NFkB, TGFb, IGF, WNT
Chronic stress signals
Collagen antibody induced;
Reactive oxygen species;
DNA damage;
Inflammation;
ER stress;
(B.)
Stem/progenitor cells
cbp/p300
RNAPII
NF-kB
Developmental genes held in a low transcription state
(I)
PRMT5
? ZFP57
KRAB
Kap1
Mbd4
Usp7
Dnmts
UHRF1
Poised transcription
SOX9, KLF2, KLF4, CDKN1A
Bivalent chromatin
(F.)
(E.)
(C.)
Certain types of cancer cells
(II)
Developmental genes transcripted in a high transcription state
Activation of stem cell pathway
Active transcription
SOX9, KLF2, KLF4, CDKN1A
Open chromatin
(D.)
Extracellular matrix, for example, upregulation of Col2a1 and Col8a1
(G.)
WNT signalling pathway, for example, downregulation of Wnt4 and Wnt5b
Defects of apoptotic signalling pathways (impaired programmed cell death) & prolonged proliferative and apoptotic responses

Figure. 5.10. A working model of possible role of MBD4 in human cancers such as breast cancers. (I.) In normal stem/progenitor cells, the promoter regions of certain developmental genes including *SOX9*, *KLF2* and *KLF4* may be marked by both active (trimethylated histone H3 lysine 4; H3K4me3) and repressive marks (trimethylated histone H3 lysine 27; H3K27me3), termed "bivalent chromatin" (Tsai and Baylin, 2011). MBD4 may recruit repressive complexes containing KAP1, PRMT5, and ZFPs as well as USP7, UHRF1, and DNMTs, which protect the binding sites at promoter regions from activation factors such as NF-kB, cbp/p300, and RNAPII to hold these genes in a low, poised transcription state. (II.) In certain types of cancer cells, inactivation and/or downregulation of MBD4 might cause deassociation of the repressive complexes. This might expose the binding sites at promoter regions to activation factors, and results in a high transcription state of these developmental genes. Meanwhile, the derepression at these promoter regions may be accompanied with open chromatin marked by active marks of H3K4me3. These schematic processes may be induced by a number of chronic stress signals including collagen antibody induced stress, reactive oxygen species, DNA damage, inflammation, and/or ER stress (A). Multiple signaling pathways including NF-kB, TGFb, EGF, and WNT may conduct these stress signals into nuclear (B). Repeated environmental stress would accumulate to a certain threshold, especially when the repression machinery has defects such as inactivation of MBD4 and/or other MeCPs and glycosylases like TDG, and/or when combined with other oncogenic background such as APC min or HRAS. This may promote the cell renewal system with genetic or epigenetic abnormalities (C). These abnormalities may include a high transcription state of certain developmental genes discussed, and the activation of stem cell pathways in these tumor cells. The consequences of such abnormalities may manifest a number of changes at gene expression level, including the upregulation of extracellular matrix genes *Col2a1* and *Col8a1*; the downregulation of WNT signaling pathway genes *Wnt4* and *Wnt5* (D). In addition, this may also cause defects of apoptotic signaling pathways (impaired programmed cell death), and prolonged proliferative and apoptotic responses, which contribute to tumor initiation and progression (D). These changes may alter the existing signaling pathways (E), and form abnormal loops to auto-activate these 'disease' processes (F), as well as lower the threshold to the repeated chronic stress signals (G). This model might also apply in the possible role of MBD4 in autoimmune diseases such as rheumatoid arthritis.

Why are *Mbd4* knockout mice viable, while MBD4 has so many important interaction proteins especially in transcriptional regulation? And why are a number of developmental and imprinted genes misregulated when MBD4 is inactivated, but only a mild phenotype was observed from these mice? My interpretation is that this might be due to the multiple transcriptional repression machineries available in the system, including other MeCPs such as MBD1, MBD2, UHRF1 and their associated co-repressors, which may well compensate the potential role of MBD4 in transcriptional repression. They might work like multiple locks that share a cohort of co-repressive components, to secure an 'optimal' expression of certain developmental genes in specific types of tissues. Lack of an individual component of the lock might only lower the threshold, but has only a mild phenotype, as these developmental genes express during early developmental stages anyway, thus their increased expressions may not cause severe phenotypes that were observed where

inactivated. In contrast, in certain types of differentiated tissues, especially at the location or tissue types that are usually exposed to chronic stresses, the upregulation of these developmental genes may result in positive feedbacks or loops to reinforce and/or amplify the systemic burden when combined on other oncogenic background, and transit the cell status and behavior to cancer proliferation or a worse prognosis (Figure. 5.10). This would be particularly the case in certain process such as epithelial-to-mesenchyme transition (EMT). Consistent with this speculation, an unpublished work demonstrated some interesting data that in gene expression level, a number of MeCPs including MBD1, MBD2, MBD4, and UHRF1, and a number of co-repressive complexes containing HDAC1, PRMT5 etc. were dramatically downregulated in transformed human mammary epithelial cells (by induce hRAS to undergo a neoplastic transformation) (Deangelis, 2010). Moreover, as mentioned in above discussion section, Sox9 has been shown to cooperatively determine the mammary stem cell state with Slug, and elevate the tumorigenic and metastasis-seeding abilities of human breast cancer cells, likely in concert with the EMT program (Guo et al., 2012).

Although it is very possible that MBD4 might act as a tumor suppressor through transcriptional repression, it does not preclude that MBD4 might also exert its influence through signaling. Actually the signaling pathways that MBD4 is involved in may link the opposing forces in certain cancers. In my gene expression study, a cohort of genes from wingless-type MMTV integration factor (WNT) signaling pathway, insulin growth factor (IGF) and insulin signaling pathway, and retinoic acid (RA) signaling pathway were dramatically misregulated in MEF cell lines that lack MBD4 activity. Interestingly, SOX9 has been shown to potentially act downstream of both WNT (in the hair follicle and intestinal crypt) (Wang et al., 2007), IGF (in the putative intestinal crypt stem cell region) (Ramocki et al., 2007) and RA (breast cancer cell lines) (Muller et al., 2009, Afonja et al., 2002) signaling pathways. However, the downstream effects of these pathways caused by MBD4 inactivation seem to differentially regulate the transcriptional factors, including SOX9, depending on the developmental stages and tissue types. For example, the WNT pathway gene *Wnt5* can differentially signal through noncanonical pathways to upregulate *Sox9* during early stages of chondrogenesis and downregulate it during chondrocyte maturation (Pritchett et al., 2011). In this context, it may warrant caution that not all the data are concordant with the paradigm model that derepression of *Sox9* by MBD4 inactivation promotes proliferation and advances certain cancer growth. For example, xenografts of LNCaP cells (Prostate cancer) containing SOX9 overexpression enhanced tumor growth (Wang et al., 2008), whereas the stably transfected M12 prostate cancer cells highly

expressing SOX9 suppressed growth and tumorigenicity (Drivdahl et al., 2004). In addition to the tissue specificity, one clue that might cause these contradictory outcomes is the differential status of a cell cycle arrest protein p21. In a report that SOX9 was observed absent from nearly all examined melanomas, overexpressing SOX9 in melanoma cell lines or in xenografts reduced cell proliferation and tumor growth, which coincided with the p21 upregulation (Passeron et al., 2009). In contrast, the lung adenocarcinoma that highly express SOX9 exhibited enhanced tumor cell proliferation, which coincided with significant lower p21 expression (Jiang et al., 2010). In addition, SOX9 was detected in over 80% of melanomas in another report (Rao et al., 2010), and I observed high nuclear expression of SOX9 as well as MBD4, and also their cytoplasmic expression (much more nuclear than cytoplasmic) in primary human breast cancer tissues (Supplementary figure 7.1). It seems that the switches of some transcriptional factors from a tumor-suppressor gene to an oncogene, depend on the status of the cell cycle arrest protein p21 (Rowland and Peeper, 2005). These include above described SOX9 (Pritchett et al., 2011) and the Kruppel-like factor KLF4. In this context, MBD4 may also act as an apoptotic signaling sensor, while derepressing and cooperating with p21 (CDKN1A) and/or KLF4, and the inactivation of MBD4 may cause cell cycle arrest and suppression of apoptosis signaling, which also facilitate its function at DNA-damage repair.

Last but not least, it is very interesting that MBD4 potentially has a direct interaction with ERα identified in my co-IP and IF, which indicate MBD4 might involve in the cellular responses under hormone stress such as Estradiol (E2 or 17β-estradiol, also oestradiol). MBD4 has been reported to be involved in transcriptional derepression of a vitamin D hormone receptor that responses to parathyroid hormone (Kim et al., 2009). Interestingly when ERα was induced and bound by estradiol and recruited to target promoters, an ordered and cyclical recruitment of DNMTs and cofactors has been reported (Métivier et al., 2008, Métivier et al., 2003). Estrogen receptor and vitamin D hormone receptor belong to the nuclear receptor superfamily of transcription factors (Chawla, 2001). This superfamily includes not only the classic endocrine receptors type I steroid hormones, type II retinoic acid receptor, retinoid X receptor and thyroid hormone receptor, and the fat-soluble vitamins A and D, but also a large cohort of 'orphan' nuclear receptors (Chawla, 2001). It is interesting that one of such 'orphan' nuclear receptors HNF4a was implicated in my MBD4 pathway analysis of both protein interactions and gene expressions (Supplementary figures 3.1 D & 5.1 F). Despite the variations of their ligand sensitivities, the structural organization of these nuclear receptors is similar (Chawla, 2001). It seems MBD4 may have a universal role in the regulation of all these classes of nuclear receptors in response to the responsive

hormone stress or inductions. This might depend on the association co-factors and/or the glycosylase activity of MBD4, which may also relevant to my model in Figure. 5.10, and need further investigation.

5.3.6. Concluding remarks

It is noteworthy that all these transcriptional aberrations by lack of MBD4 were observed in MEFs, a mixture of different types of cells and before end of developmental stage. Considering the mild phenotype of *Mbd4* knockout mice, the significant phenotype caused by lack of MBD4, however, might emerge in differentiated tissues and at a late onset or under certain stress. Thus it is required to choose more relevant cells or tissues to validate the possible roles of Mbd4 following these clues. In addition, it will be of interest to investigate the expression level and pattern of MBD4 in line with its possible targets identified such as *Sox9*, *Klf2*, *Klf4*, and the possible compensators such as TDG (and/or other MeCPs) in terms of its roles in transcriptional repression and/or glycosylase activity, in some relevant normal and diseased tissues including breast and synovium. Lastly, MBD4 rescue experiments are essentially necessary to identify the direct gene targets of MBD4 for further investigations.

Chapter 6: Evolutionary analysis of in vitro binding specificity of MBD domain of MBD4.

6.1. Introduction

It has been previously determined that the MBD domain of human and mouse MBD4 can recognise and bind to single symmetrically methylated CpG dinucleotides in test substrates in vitro (Hendrich and Bird, 1998, Bellacosa et al., 1999). In addition, the G:T mismatches derived from deamination of methylated CpG dinucleotides can be both bound by the N terminus of MBD4 and processed by its glycosylase domain in vitro (Hendrich et al., 1999). However, it is unknown if the binding specificity of MBD domain of MBD4 is conserved between species in vertebrate. While the phenotype of *Mbd4* knockout in mice is viable, the *Mbd4* morpholino knockdown in frog results in late lethal phenotype, and they fail to develop from tadpole to adult frog (Ruzov et al., 2009), indicating a central role of MBD4 in the early development of lower vertebrates. However, a pertinent question would be if MBD4 in lower vertebrates like frog and/or fish also possesses a similar binding specificity to methylated CpG dinucleotides.

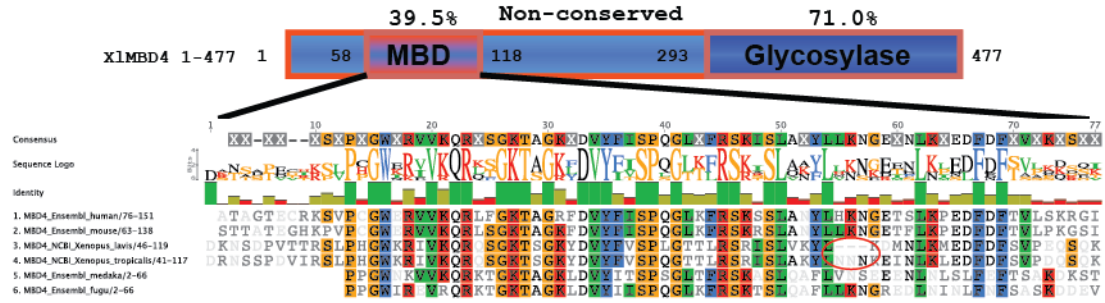
In general the MBD domain of MBD4 is most similar to that of MeCP2 within the methy-CpG binding domain proteins (Hendrich and Tweedie, 2003). The glycosylase domain of MBD4 is very conserved between species (77.9% pairwise identical primary sequences between human and Xenopus, 79.6% between human and Medaka fish). However, unlike MeCP2 whose MBD domain is very conserved between species (94.7% pairwise identical primary sequences between human and Xenopus, 86.8% between human and Medaka fish) (Figure. 6.1.1. B), the MBD domain of MBD4 is poorly conserved (44.7% pairwise identical primary sequences between human and Xenopus, 49.2% between human and Medaka fish, in contrast to 46.1% pairwise identical primary sequences between human MBD4 and MeCP2) (Figure. 6.1.1. A). The primary sequences within the MBD domain are important to the binding specificities of methyl-CpG binding proteins (MeCPs). One example is that the mammalian MBD3 protein has a very complete MBD domain, which is very similar to that of MBD1 and MBD2, but it was poor recognition of methylated DNA due to point substitutions in its MBD domain (Ballestar and Wolffe, 2001). The solution structures of MBD domains of MeCP2 and MBD1 have been solved (Ohki et al., 1999, Wakefield et al., 1999, Ohki et al., 2001) (Figure. 6.1.2 structure and loops indicated). Key amino acids responsible for their methyl-CpG recognition, DNA binding conservation, interaction with DNA bases, MeCP2 mutations, and structure and loop have been identified (Free et al., 2001, Ohki et al., 1999, Wakefield et al., 1999, Ohki et al., 2001) (Figure. 6.1.2 red dots above). These key functional and structural amino acids are well conserved within MBD domain of MBD4, despite its poorly conserved primary sequences between species

(Figure. 6.1.2 The lanes where grey vertical lines indicated). Thus it is very possible that MBD domains of MBD4 in lower vertebrates like Xenopus frog and Medaka fish still possess the binding specificity to the methylation marks despite of their poor conserved primary sequences within MBD domain.

Given their common MBD domains, different MeCPs may possess special binding specificities of the context of upstream or downstream signature sequences in addition to the methyl-CpGs. For example, MeCP2 cannot occupy the loci vacated by MBD2 (Klose et al., 2005). In addition, in two cancer cell lines of MCF7 and MDA- MB-231, while many genes are associated with multiple MeCPs, there are a number of genes regulated only by MeCP2, other than MBD1 or MBD2 (Ballestar et al., 2003), such that ~9% of MeCP2 binding sites were not IPed by MBD1 or MBD2 chromatin immunoprecitation (ChIP) (Klose et al., 2005). These suggest there might be recognition differences between MeCP2 and MBD1 or MBD2 in addition to their common object; methyl-CpG. MeCP2 is the closest relative of MBD4 within the group, while other MeCPs may share another common ancestor (Albalat, 2008). The family tree of the MBD domain of MeCPs supports the hypothesis that in addition to the necessary methyl-CpG, different MeCPs might have distinct binding ability to some special context sequences. Their intrinsic primary sequences within or outside of MBD domain might be responsible for such binding specificity. For example, MeCP2 may require the ~13 amino acids upstream in addition to MBD domain for its high-affinity binding to the special DNA bases of A/T runs adjacent to methyl-CpG (Klose et al., 2005).

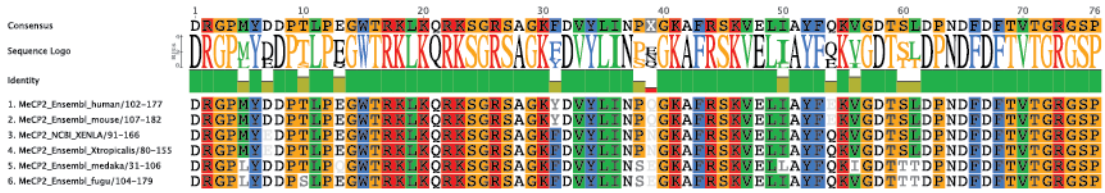
DNA bases in the CpG context in cellular environment are vulnerable to chemical reactions. One of the most frequent DNA damages affecting the base-pairing properties at CpG dinucleotides is the hydrolytic deamination of cytosine and 5-methylcytosine (5-meC), which results in uracil and thymine mispaired with guanine, the G:U and G:T mismatches respectively (Jacobs and Schär, 2011). If these mismatches are not repaired before DNA replication, the cells would acquiesce them and it may cause C:G to T:A transition mutations. In mammals, MBD4 is one of the four uracil glycosylases, and one of only two thymine glycosylases in the system (Jacobs and Schär, 2011). Unlike other glycosylases, it has a MBD domain that potentially recognise deamination products at CpG context in a specific manner. While it has been shown that MBD4 can process a wide range of G-mispaired base lesions including thymine, uracil, 5-FU, and 3,N4-ethenocytosine (Jacobs and Schär, 2011, Petronzelli et al., 2000), the binding affinities of the MBD domain of MBD4 to deamination mismatches in different methylation contexts such as opposite to unmethylated, fully methylated or hemimethylated CpGs are not clear. In addition, all the other uracil glycosylases lack the ability to specifically recognise the CpG dinucleotides,

A



	hMBD4	mMbd4	XlMBD4	XtMBD4	MedMBD4	FuguMBD4
hMBD4	-	81.6	44.7	50.0	49.2	53.8
mMbd4		-	46.1	51.3	52.3	56.9
XlMBD4			-	81.8	43.1	41.5
XtMBD4				-	50.8	44.6
MedMBD4					-	64.6
FuguMBD4						-

B



	MBD4	Mbd4	XlMBD4	XtMBD4	MedMBD4	FuguMBD4
hMeCP2	-	100	94.7	94.7	86.8	89.5
mMeCP2		-	94.7	94.7	86.8	89.5
XlMeCP2			-	100	88.2	90.8
XtMeCP2				-	88.2	90.8
MedMeCP2					-	94.7
FuguMeCP2						-

In A

1. MBD4_Ensembl_human/76-151
2. MBD4_Ensembl_mouse/63-138
3. MBD4_NCBI_Xenopus_lavis/46-119
4. MBD4_NCBI_Xenopus_tropicalis/41-117
5. MBD4_Ensembl_medaka/2-66
6. MBD4_Ensembl_fugu/2-66

In B

1. MeCP2_Ensembl_human/102-177
2. MeCP2_Ensembl_mouse/107-182
3. MeCP2_NCBI_Xenopus_lavis/91-166
4. MeCP2_Ensembl_Xenopus_tropicalis/80-155
5. MeCP2_Ensembl_medaka/31-106
6. MeCP2_Ensembl_fugu/104-179

Figure. 6.1.1 The domain structure of MBD4 vary in their sequence conservation. A. Schematic domain structure of *xenopus lavis* Mbd4 is shown, with amino acids indicated showing the beginning and end of the significant domains. The percentage above the domain is the conservations between the species of human and stickleback Mbd4. The alignment of MBD domain of MBD4 is shown. The red circle indicates 3 amino acids missing in the primary sequence of *xenopus lavis* Mbd4. The table shows the conservation between the selected species for the MBD domain of MBD4. The numbers in the table are the percentage of pairwise identical sites between the two indicated species. B. The alignment and conservation of primary sequences of MeCP2. The MBD from MeCP2 is more highly conserved. But note that the MBD domains of MeCP2 and MBD4 are most conserved between MBD proteins (cf MBD1, MBD2 and MBD3). The species in the alignments are human sapiens (h); *Mus musculus* (m); *Xenopus lavis* (Xl); *Xenopus tropicalis* (Xt); *Oryzias latipes* (Med); *Takifugu rubripes* (Fugu).

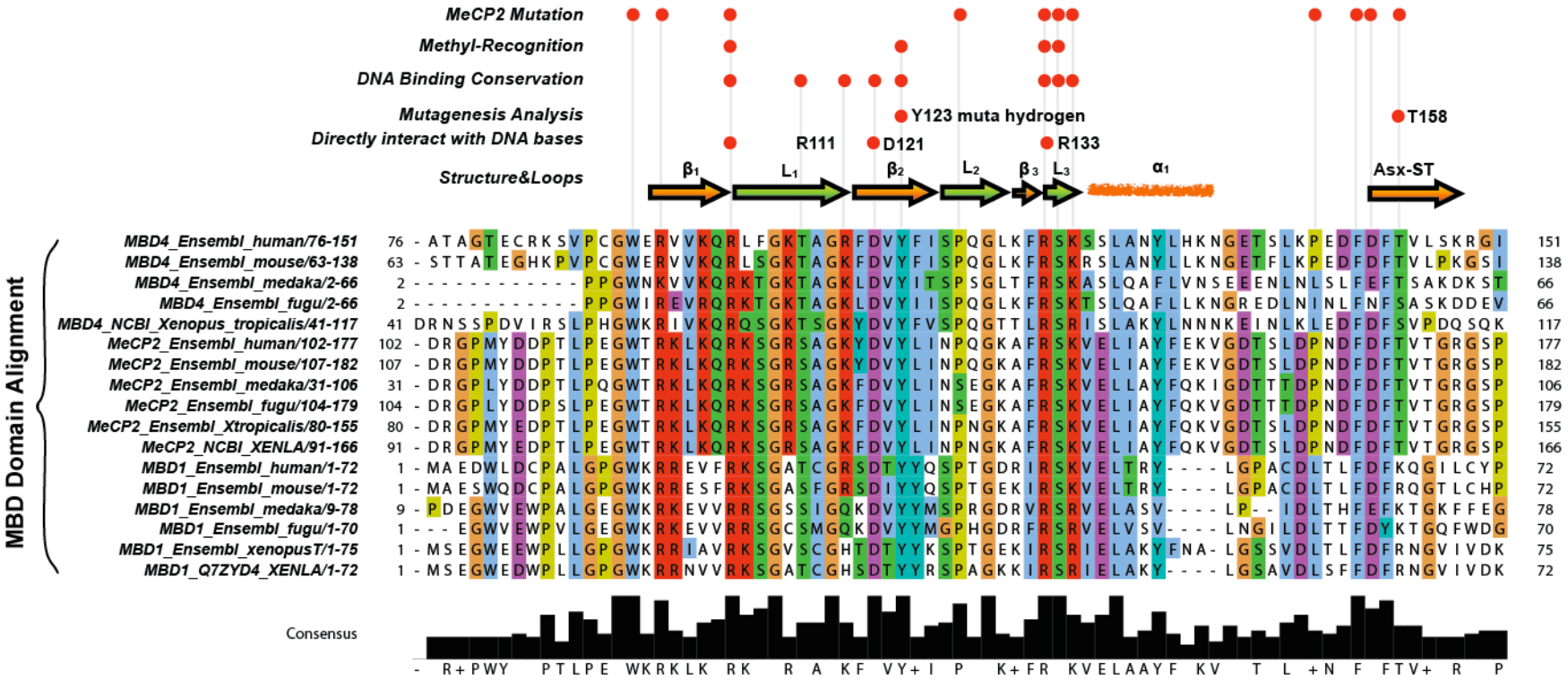

MeCP2 Mutation
Methyl-Recognition
DNA Binding Conservation
Mutagenesis Analysis
Directly interact with DNA bases
Structure&Loops
Y123 muta hydrogen
T158
R111
D121
R133
β1 L1 β2 L2 β3 L3 α1
Asx-ST
MBD Domain Alignment
MBD4_Ensembl_human/76-151 76 - ATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGI 151
MBD4_Ensembl_mouse/63-138 63 - STTATEGHKPVPCGWERVVKQRLSGKTAGKFDVYFISPQGLKFRSKRSLANYLLKNGETFLKPEDFDFTVLPKGSI 138
MBD4_Ensembl_medaka/2-66 2 - - - - - - - - - - - - PPGWNKVVKQRKTGKTAGKLDVYITSPSGLTFRSKASLQAFLVNSEEENLNLSLFEFTSAKDKST 66
MBD4_Ensembl_fugu/2-66 2 - - - - - - - - - - - - PPGWIREVRQRKTGKTAGKLDVYIISPQGLKFRSKTSLQAFLLKNGREDLNINLFNFSASKDDEV 66
MBD4_NCBI_Xenopus_tropicalis/41-117 41 DRNSSPDVIRSLPHGWKRIVKQRQSGKTSGKYDVYFVSPQGTTLRSRISLAKYLNNNKEINLKLEDFDFSVPDQSQK 117
MeCP2_Ensembl_human/102-177 102 - DRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSP 177
MeCP2_Ensembl_mouse/107-182 107 - DRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSP 182
MeCP2_Ensembl_medaka/31-106 31 - DRGPLYDDPTLPQGWTRKLKQRKSGRSAGKFDVYLINSEGKAFRSKVELLAYFQKIGDTTTDPNDFDFTVTGRGSP 106
MeCP2_Ensembl_fugu/104-179 104 - DRGPLYDDPSLPEGWTRKLKQRKSGRSAGKFDVYLINSEGKAFRSKVELIAYFQKVGDTTTDPNDFDFTVTGRGSP 179
MeCP2_Ensembl_Xtropicalis/80-155 80 - DRGPMYEDPTLPEGWTRKLKQRKSGRSAGKFDVYLINPNGKAFRSKVELIAYFQKVGDTSLDPNDFDFTVTGRGSP 155
MeCP2_NCBI_XENLA/91-166 91 - DRGPMYEDPTLPEGWTRKLKQRKSGRSAGKFDVYLINPNGKAFRSKVELIAYFQKVGDTSLDPNDFDFTVTGRGSP 166
MBD1_Ensembl_human/1-72 1 - MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRY- - - - LGPACDLTLFDFKQGILCYP 72
MBD1_Ensembl_mouse/1-72 1 - MAESWQDCPALGPGWKRRESFRKSGASFGRSDIYYQSPTGEKIRSKVELTRY- - - - LGPACDLTLFDFRQGTLCHP 72
MBD1_Ensembl_medaka/9-78 9 - PDEGWVEWPALGEGWKRKEVVRRSGSSIGQKDVYYMSPRGDRVRSRVELASV- - - - LP- - IDLTHFEFKTGKFFEG 78
MBD1_Ensembl_fugu/1-70 1 - - - EGWVEWPVLGEGWKRKEVVRRSGCSMGQKDVYYMGPHGDRFRSRVELVSV- - - - LNGILDLTTFDYKTGQFWDG 70
MBD1_Ensembl_xenopusT/1-75 1 - MSEGWEEWPLLGPGWKRRIAVRKSGVSCGHTDTYYKSPTGEKIRSRIELAKYFNA- LGSSVDLTLFDFRNGVIVDK 75
MBD1_Q7ZYD4_XENLA/1-72 1 - MSEGWEDWPLLGPGWKRRNVVRKSGATCGHSDTYYRSPAGKKIRSRIELAKY- - - - LGSAVDLSFFDFRNGVIVDK 72
Consensus
- R+PWY PTLPE WKRKLK RK R A KF VY+I P K+FR KVELAAYF KV T L +N F FTV+ R P

Figure. 6.1.2 Key functional and structural amino acids are well conserved within MBD domain of MBD4. The MBD domain alignment is shown. A solution structure of the MBDs from MeCP2 and MBD1 has been determined, consisting of four anti-parallel β-strands, two of which were proposed to interact with the major groove of DNA, where a methyl group would be located. In addition, a number of conserved residues throughout the MBD domains of MeCP2 and MBD1 can be easily revealed by alignment, despite their full-length sequences sharing only moderate homology. The MBDs of MBD4, MBD1 and MeCP2 were aligned and compared to indicate essential residues within the MBD of MBD4 that are responsible for binding to methylated DNA sequences. Essential Residues are well conserved in the MBD of MBD4. In addition, amino acids within the MBD of MBD4 that are important for DNA binding found by mutational analyses and associated with Rett syndrome in MeCP2 are also well conserved. Sequence alignments were generated with clustalX module of Jalview software.

while such promoter associated sites conceivably need careful surveillance, implicating possible unknown machinery to specifically distinguish this deamination mismatches at CpG sites from the base damages of vast genomes.

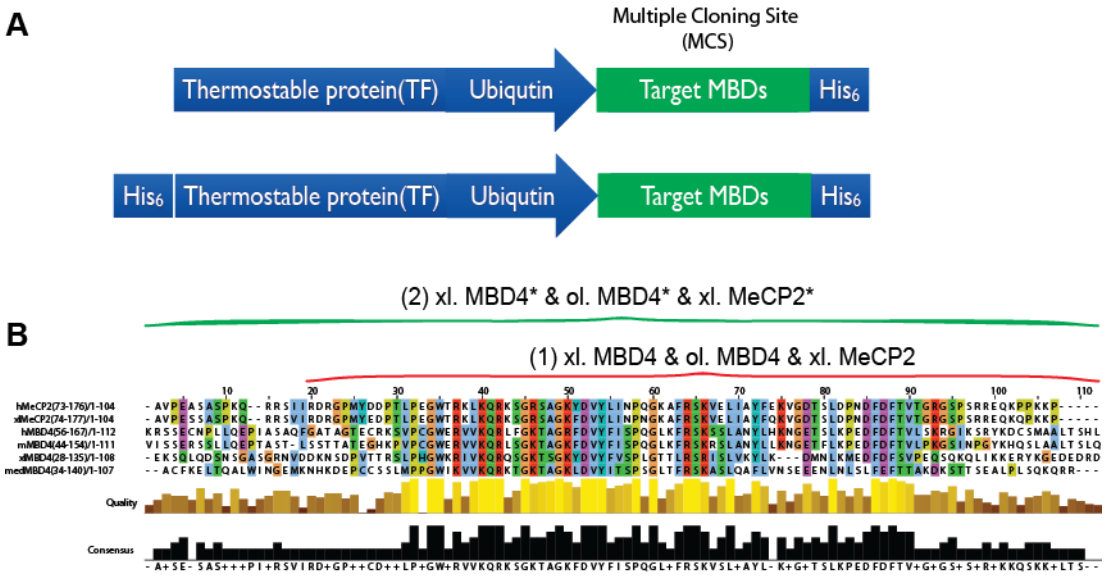
In this chapter, I first performed a comparison of MBD domain DNA binding activity with distant MBD4 homologues from the Medaka fish (*Oryzias latipes*) and the amphibian, *Xenopus laevis* to probes containing 5mCpG. I then checked the DNA binding specificity of MBD4 and MeCP2 to the probes containing the newly discovered hydroxymethylated mark. Lastly, their binding affinities to the G:U mismatches in a serial of unmethylated, fully methylated and hemimethylated context were tested. The surprising results suggest a previous unknown contribution to efficient repair of this deamination product at CpG dinucleotides by MeCP2, as well as a well conserved binding specificity of MBD4 to 5mCpG dinucleotides.

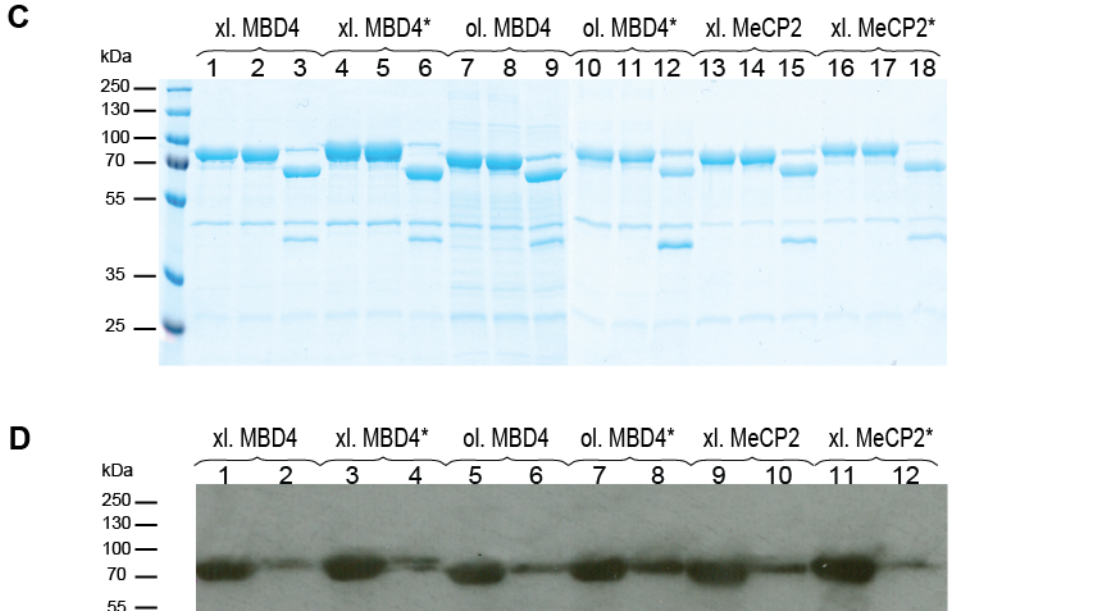
6.2. Results

6.2.1. Generation of 6HIS- tagged TF-UB-MBD fusion proteins

I chose a vector system called 6HIS-TF-UB (Figure. 6.2. A) for productions of MBD domain peptides of Mbd4 homologues from the Medaka fish (*Oryzias latipes*) and the amphibian frog (*Xenopus laevis*) (Figure. 6.2 B). This vector system contains *Escherichia coli* thermostable protein trigger factor (TF), which shows high expression and stability, and is capable of helping the proteins cloned next to TF express to be in a soluble state (Thapa et al., 2008). It also engineered the minimum ubiquitin sequences downstream of TF, linking to the targeting proteins, such that the TF-UB fusion proteins can be cleavaged by specific

ubiquitin cleavage enzymes (Thapa et al., 2008, Baker et al., 2005) (Figure. 6.2. A). However, the empty TF-UB vector I received from the Park lab (Thapa et al., 2008) only has a 6HIS tag at its downstream (Figure. 6.2. A upper), and there is a stop codon within its multiple cloning site (MCS), which results in a control TF-UB protein without the 6HIS tag. In addition, the fusion proteins containing the target MBDs cannot be observed after cleavage under western blotting. Thus I decided to insert another 6HIS tag upstream of the TF by site-directed mutagenesis (the QuikChange® II XL Site-Directed Mutagenesis Kit (Stratagene)) (see Chapter 2. Method) (Figure. 6.2. A lower). I then cloned the MBD domain of Mbd4 of *Xenopus laevis* (xl.Mbd4), *Oryzias latipes* (ol.Mbd4) and MBD domain of MeCP2 of *Xenopus laevis* (xl.MeCP2), into the 6HIS-TF-UB system (Figure. 6.2 B). Specifically, amino acids 49-134 of ol.Mbd4, 44-135 of xl.Mbd4, and 90-177 of xl.MeCP2 (Figure. 6.2. B (1) red bracket) were cloned downstream of TF-UB, followed by 6his-tag at their C-terminal. In addition, I also made another batch of these proteins, which included the ~13 amino acids upstream of their MBD domains (Figure. 6.2. B (2) green bracket, labeled *), which were suggested to confer high-affinity binding of MeCP2 to the special DNA bases of A/T runs adjacent to methyl-CpG (Klose et al., 2005). These fusion proteins were produced in *E.coli*, and then purified by HIS-tagged Ni-NTA columns (see methods). The fusion MBD proteins were very soluble, and can be cleaved by ubiquitin specific enzyme USP2cc (Baker et al., 2005), suggesting they are suitable for the following EMSA assays (Figure. 6.2. C&D).


A
Multiple Cloning Site
(MCS)
Thermostable protein(TF)
Ubiqutin
Target MBDs
His6
His6
Thermostable protein(TF)
Ubiqutin
Target MBDs
His6
B
(2) xl. MBD4* & ol. MBD4* & xl. MeCP2*
(1) xl. MBD4 & ol. MBD4 & xl. MeCP2
Quality
Consensus


C
xl. MBD4
xl. MBD4*
ol. MBD4
ol. MBD4*
xl. MeCP2
xl. MeCP2*
kDa
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
250
130
100
70
55
35
25

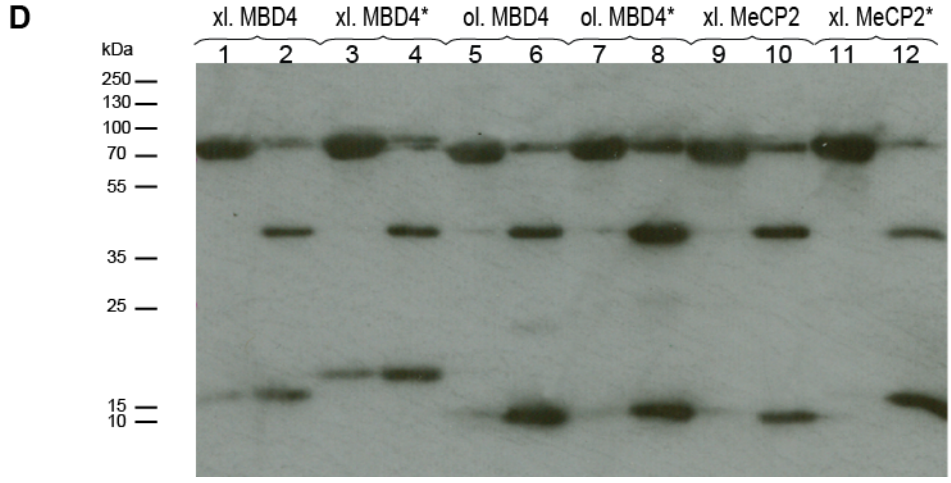

D
xl. MBD4
xl. MBD4*
ol. MBD4
ol. MBD4*
xl. MeCP2
xl. MeCP2*
kDa
1 2 3 4 5 6 7 8 9 10 11 12
250
130
100
70
55
35
25
15
10

Figure. 6.2. Generation of 6HIS- tagged pET21b-TF-UB-MBDs fusion proteins. (A.) The schematic representation of the fusion protein of MBD domain of MBD4. (B.) The alignment of MBD domain of MBD4 and MeCP2 of human (h), mouse (m), the Medaka fish (*Oryzias latipes*, labeled med or ol) and the amphibian, *Xenopus laevis* (labeled x or xl). The differences of ~ 20 amino acids (AAs) at upstream between the green bracket (1) and the red bracket (2) are those reported crucial AAs for specific binding to mCpG with A/T run in hMeCP2 (Klose et al., 2005). The fusion proteins with these ~20 AAs are labeled with *. (C-D) Cleavage of recombinant proteins expressed in *E. coli* at ubiquitin site. (C.) SDS–PAGE analysis of protein fractions with and without cleavage. The corresponding fusion proteins were indicated with bracket; Lane 1, 4, 7, 10, 13, 16 were purified soluble fusion proteins; lane 2, 5, 8, 11, 14, 17 were same proteins treated at 37°C for 1 hour without cleavage protein enzyme; lane 3, 6, 9, 12, 15, 18 were reactions where excess ubiquitin specific cleavage protein enzyme, 6HIS-Usp2-cc was added and incubated at 37°C for 1 hour. (D) Western Blotting (WB) analysis of protein fractions with and without cleavage. Lane 1, 3, 5, 7, 9, 11 were purified soluble fusion proteins; lane 2, 4, 6, 8, 10, 12 were reactions where excess ubiquitin specific cleavage protein enzyme, 6HIS-Usp2-cc was added and incubated at 37°C for 1 hour.

6.2.2. The binding specificity of MBD4 to methylated probes is well conserved between species, regardless of its poor sequence conservation

The MBD fusion proteins were assayed for their binding affinities to methylated and unmethylated 39-mer duplex oligonucleotides containing a single, symmetrically methylated CpG dinucleotide (Figure. 6.3. labeled color green) with A/T runs (Figure. 6.3. labeled color orange) adjacent to the CpG dinucleotide (Probe S1 and S2) at increment amount of respective proteins and fixed amount of DNA probe concentrations, and I optimised the assays with and without competitors, to ensure the binding specificities were specific (see method). Both fusion proteins of MBD domains of ol.Mbd4 and xl.Mbd4 exhibited a range of binding preferences to methylated probes (Figure. 6.3. A&B left) compared to unmethylated probes under the experimental conditions (Figure. 6.3. A&B right). In addition, the binding specificities of the MBD domain of ol.Mbd4 and xl.Mbd4 to the methylated probes with adjacent AT runs are similar to those of their responsive fusion proteins that have additional upstream ~13 amino acids (AAs) (Figure. 6.3 A&B left ol.MBD4vs*ol.MBD4 & xl.MBD4vs*xl.MBD4), suggesting in this condition, these upstream AAs do not affect the binding affinities of MBD domains of ol.Mbd4 and xl.Mbd4 to the probes containing the methylated CpG with adjacent A/T runs. It is noteworthy that I did not observe the binding specificities of the MBD domain of ol.MeCP2 and xl.MeCP2 to the methylated probes with adjacent AT runs described in Klose's report (Klose et al., 2005) under my experimental condition (Supple Figure. 6.1). In addition, my xl.MeCP2 protein

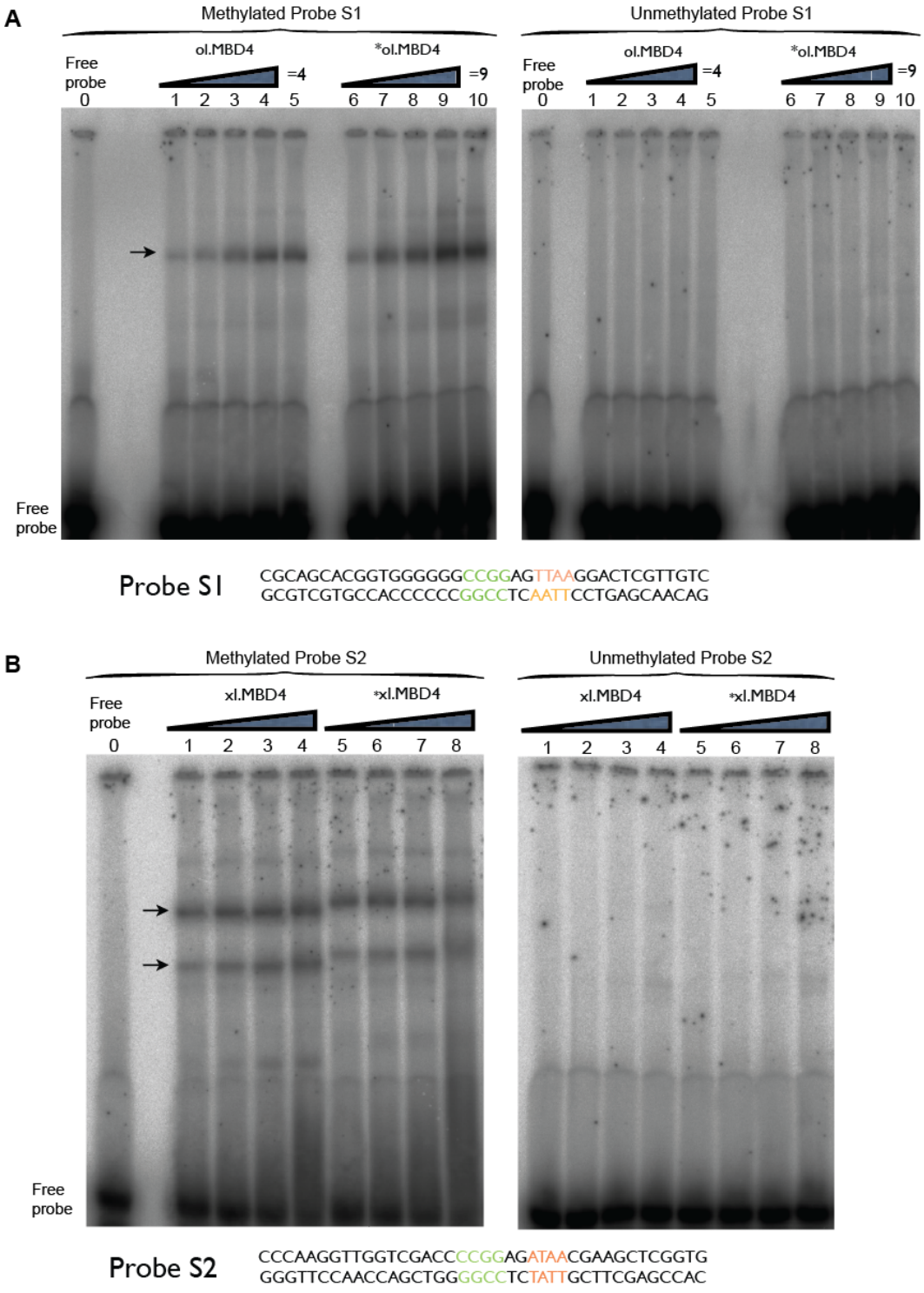

A
Methylated Probe S1
Unmethylated Probe S1
Free probe
ol.MBD4
*ol.MBD4
=4
=9
0 1 2 3 4 5 6 7 8 9 10
Free probe
Probe S1
CGCAGCACGGTGGGGGGCCGGAGTTAAGGACTCGTTGTC
GCGTCGTGCCACCCCCCGGCCTCAATTCCTGAGCAACAG
B
Methylated Probe S2
Unmethylated Probe S2
Free probe
xl.MBD4
*xl.MBD4
0 1 2 3 4 5 6 7 8
Free probe
Probe S2
CCCAAGGTTGGTCGACCCCGGAGATAACGAAGCTCGGTG
GGGTTCCAACCAGCTGGGGCCTCTATTGCTTCGAGCCAC

Figure. 6.3. The binding specificity of MBD4 with methylated probes is well conserved between species, regardless of its poor sequence conservation. Binding of the duplex probes containing 5-methylcytosine (5mC)(left) or the responsive unmethylated cytosines (right) in the context of CpG to varying concentrations of the methyl-binding domain (MBD) of (A. ol.- *Oryzias latipes*) and (B. xl- *Xenopus laevis*) MBD4 from 100 to 800nM; sequence of duplex is seen in the Figure. The correct bands are indicated by black arrows. The procedure of radioactive EMSAs is in methods.

without the ~13aa upstream has very clear binding specificity to the U6 probe (Figure. 2.1), which doesn't contain the AT runs and was claim to have no detectable binding with MeCP2 in Klose's report (Klose et al., 2005), suggesting at least in *xenopus lavis* and under my experimental condition, the ~13 amino acids (AAs) upstream of MeCP2 and the AT runs downstream of the probes do not affect the binding specificity of MeCP2.

6.2.3. ol.Mbd4, xl.Mbd4 and xl.MeCP2 have no binding specificity with hydroxymethylated probes by themselves.

Among the methylated-CpG binding domain (MBD) family of proteins, MBD4 has affinity to mCpG sites with a subsequent ability for glycosylase repair. 5-hydroxymethylcytosines (5hmCs) are the oxidative products of 5-methylcytosines recently suggested as a potential substrate in the context of active DNA demethylation (Wu and Zhang, 2011). Thus I proceeded to test the binding affinities of MBD domains of xl.Mbd4 and xl.MeCP2 to 5hmC-containing probes. Using the fusion 6HIS-tagged TF-UB-MBD of ol.Mbd4 and xl.Mbd4 proteins, I tested the affinity of these proteins to fluorescent-end-labeled 36-mers probe S2CpG containing unmethylated, methylated or hydroxymethylated cytosines at two separated CpG sites (Figure. 6.4). EMSA assays using a 6% non-denaturing polyacrylamide gel were performed. As seen in Figure 6.4 A&B&C, all of these fusion proteins can only bind to methylated but none of them bind to unmethylated or hydroxymethylated probes by themselves in this condition.

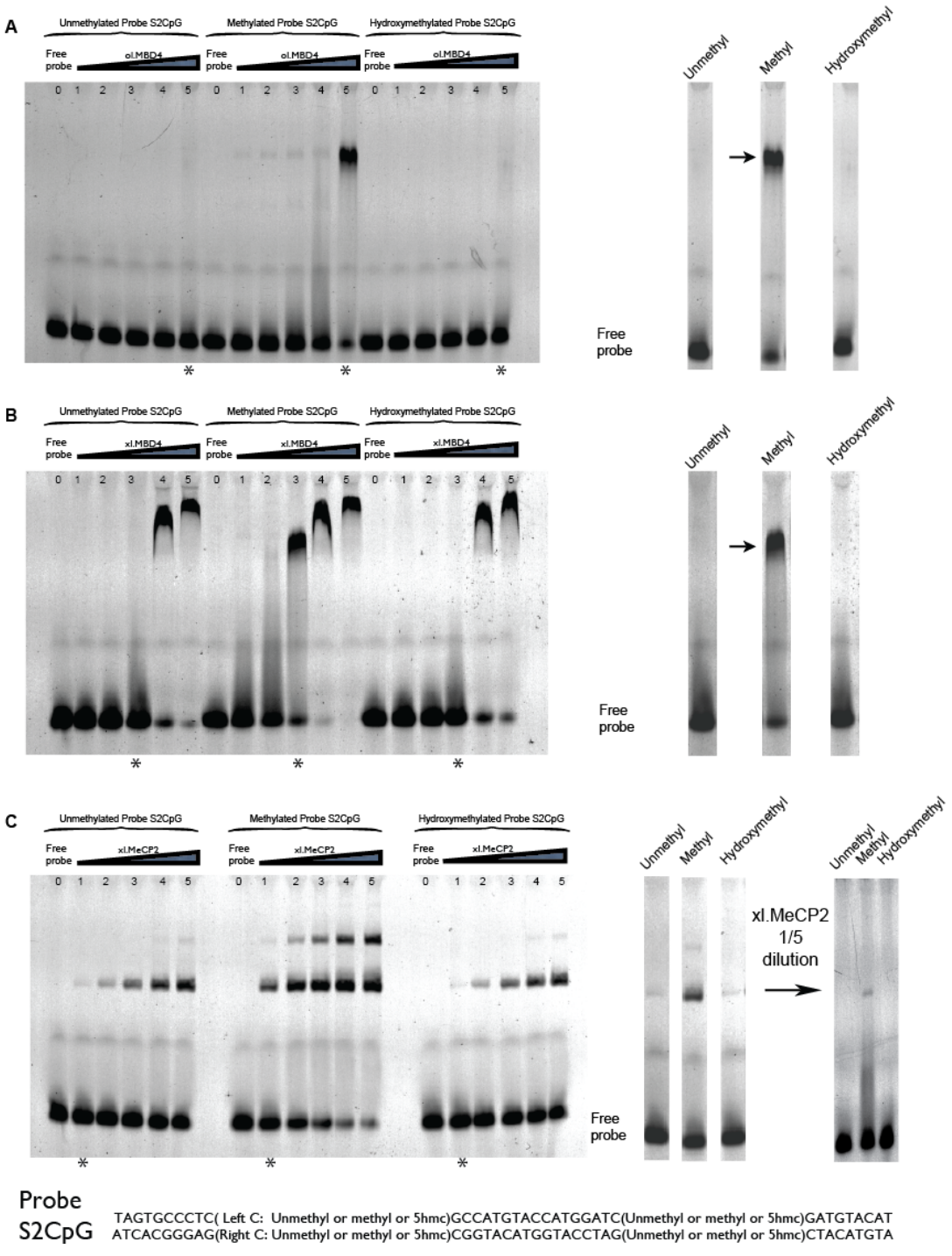
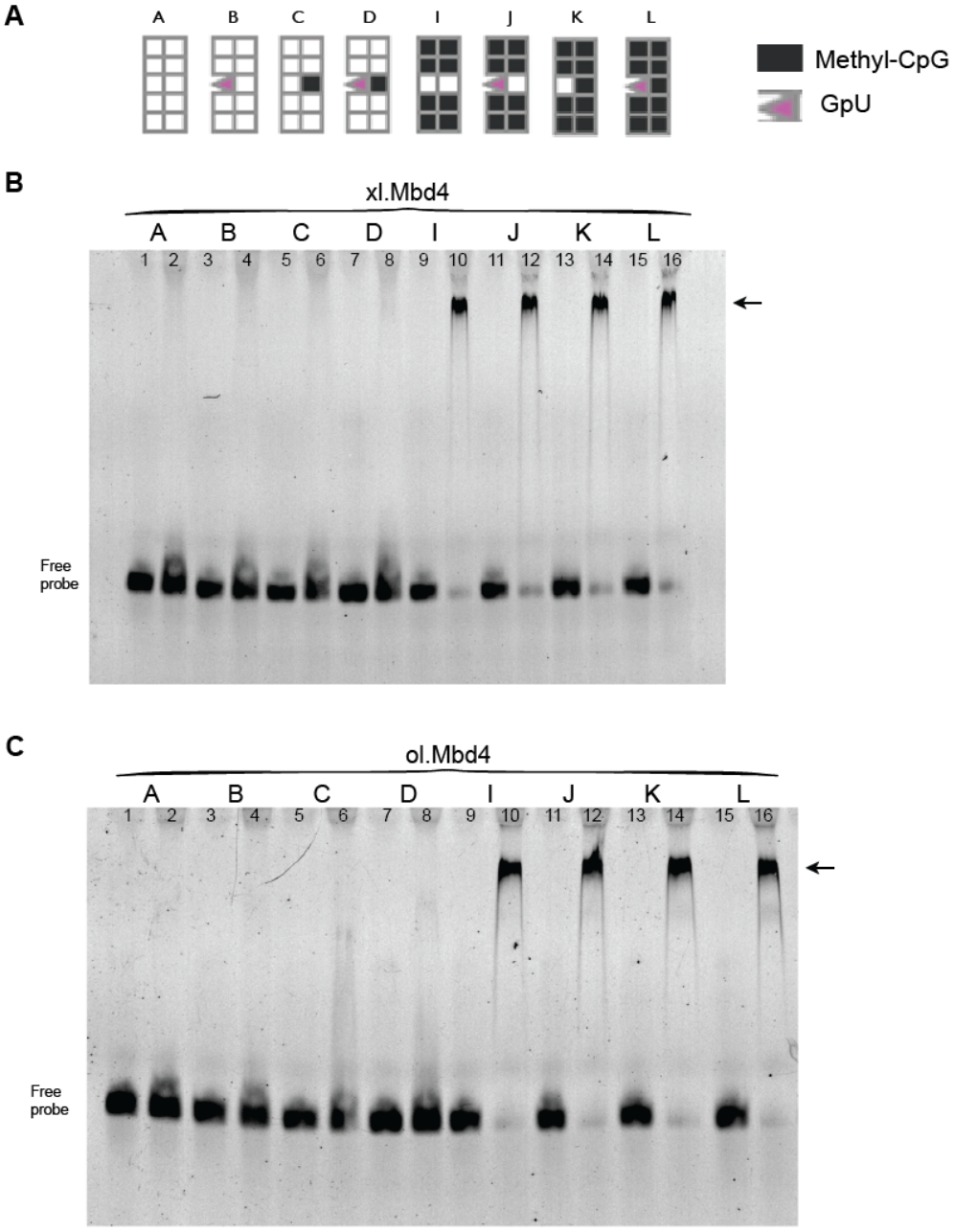


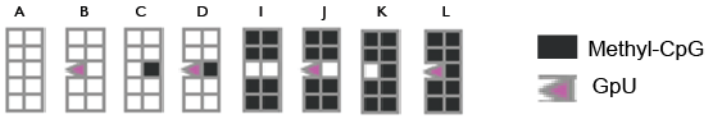
Figure. 6.4. ol.Mbd4, xl.Mbd4 and xl.MeCP2 have no binding specificity with hydroxymethylated probes. Affinity of MBD domains of ol.Mbd4 (A), xl.Mbd4 (B), and xl.MeCP2 (C) towards unmethylated cytosine, 5mC- and 5hmC-containing oligomers are shown. These proteins by themselves bind effectively to methylated CpG sequences but do not bind to the same sequences containing 5hmC in place of 5mC. The lanes with a * below on the left are corresponsive to the separated lanes on the right. The procedure of fluorescent EMSAs is in methods.

6.2.4. ol.Mbd4 and xl.Mbd4 can not distinguish the uracil mismatches at CpG dinucleotide context. In contrast, xl.MeCP2 shows a significant binding preference to semi-methylated probe with uracil mismatches.

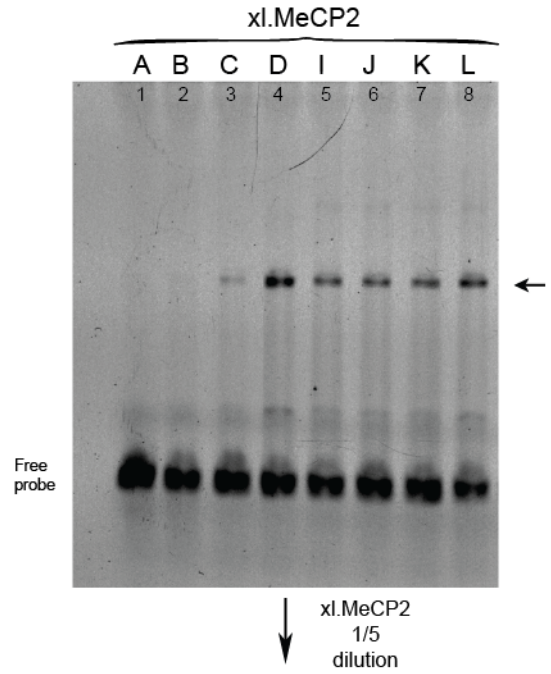
MBD4 is one of the glycosylases that can excise the uracil mismatches in vitro. However, its binding affinity to G:U mismatches in unmethylated, fully methylated, and hemimethylated context is not clear. Pierre-Antoine Defossez lab at CNRS of Universite Paris-Diderot generated a series of G:U probes. These probes are double-stranded oligonucleotides that have five separated CpGs. One of the CpG dinucleotides in the center is either unmethylated (Figure. 6.5. A. Probes A.C.I.K.), or mutated to G:U mismatches (Probes B.D.J.L.), which are opposite to an unmethylated (Probes A.B.I.J.) or methylated CpG (Probes C.D.K.L.). The other four CpG dinucleotides are either unmethylated (Probes A.B.C.D.), or fully methylated (Probes I.J.K.L.). I tested the binding affinity of MBD domain of xl.Mbd4 and ol.Mbd4 to these G:U containing probes. Not surprisingly, both fusion proteins bind to the probes containing four fully methylated symmetrically CpGs (Figure. 6.5. B&C probes I.J.K.L), while none of them bind to the responsive unmethylated ones (Figure. 6.5. B&C probes A.B.) or the probes containing one hemimethylated CpG (Figure. 6.5. B&C probes C.D.). However, both ol.Mbd4 and xl.Mbd4 do not bind the G:U mismatches probes with greater affinity (Figure. 6.5. B&C).

Surprisingly, the MBD domain of xl.MeCP2 showed very strong binding preference to the probe containing one G:U mismatch opposite to a methylated CpG (Figure. 6.5. D. Probe CvsD & KvsL). The binding affinity of xl.MeCP2 is approximately ~30-~50 fold higher than that of xl.Mbd4, and ~70-~100 fold than that of ol.Mbd4, based on my estimation of their binding affinity to the same methylated probe by their fusion protein titrations. The G:U binding affinity of xl.MeCP2 is even much stronger than its binding to the fully methylated probes (Figure. 6.5 D. Probe D vs I.J.K.L.), and the methylation at adjacent CpGs inhibited such binding affinity (Figure. 6.5 D. Probe D vs L), suggesting a unique role of MeCP2 to recognise such special G:U mismatches, especially in a context of hemimethylated G:U at a single symmetrical CpG where adjacent CpGs are unmethylated.


A
A B C D I J K L
Methyl-CpG
GpU
B
xl.Mbd4
A B C D I J K L
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Free probe
C
ol.Mbd4
A B C D I J K L
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Free probe


A
B
C
D
I
J
K
L
Methyl-CpG
GpU

D


xl.MeCP2
A B C D I J K L
1 2 3 4 5 6 7 8
Free probe
xl.MeCP2 1/5 dilution

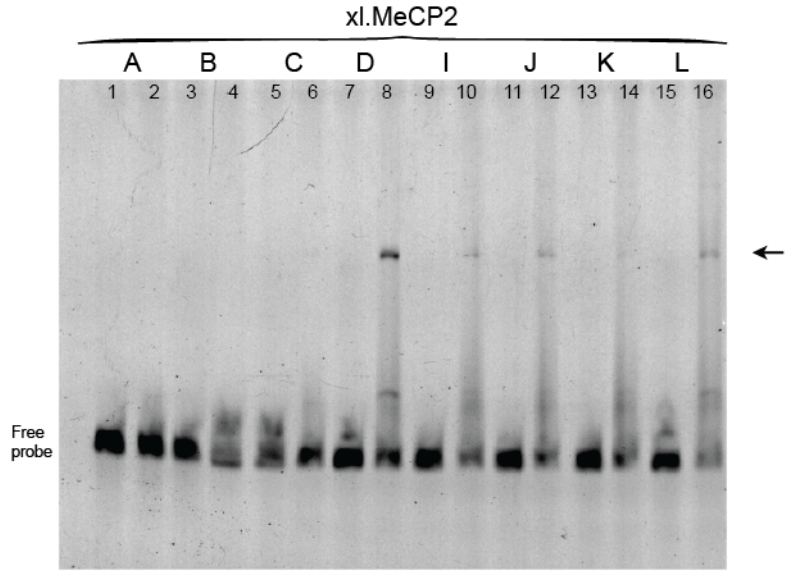

xl.MeCP2
A B C D I J K L
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Free probe

Figure. 6.5. ol.Mbd4 and xl.Mbd4 can not distinguish the uracil mismatches at CpG dinucleotide context. In contrast, xl.MeCP2 shows a significant binding preference to hemi-methylated probe with uracil mismatches. (A.) The schematic representation of the G:U probes. The fluorescent probes are 60 mer with 5 CpGs. The CpG dinucleotides are unmethylated (white square), methylated (black square), or substituted with a uracil mismatch (pink triangle). (B-C) The MBD domains of ol.Mbd4 and xl.Mbd4 bind to fully methylated probes (I, J, K, L), but can not distinguish the uracil mismatches at CpG dinucleotide context. (D.) The MBD domain of xl.MeCP2 preferentially binds to hemi-methylated probe with uracil mismatches (probe D and L). The odd number lanes in (B. C. and lower D) are free probes only and have no proteins added.

6.3. Discussion

In this chapter I have investigated the binding affinities of ol.Mbd4, xl.Mbd4 and xl.MeCP2 to a variety of probes containing modified bases at CpG sites. In sum, the binding specificity of MBD4 to methylated CpG dinucleotides is well conserved in lower vertebrates like medaka fish and xenopus frog. In addition, none of my fusion proteins of MBD domains of MBD4 and MeCP2 by themselves showed binding affinity to the probes containing hydroxymethylated CpGs. Lastly, the MBD domains of ol.Mbd4 and xl.Mbd4 did not distinguish the probes containing G:U, whereas the MBD domain of xl.MeCP2 showed a striking binding affinity to the G:U mismatch at a single hemimethylated CpG context.

6.3.1. The binding conservation of MBD4 suggests a general machinery in terms of MBD4 function during vertebrate evolution

I started my PhD from the EMSA assays to ask if MBD4 in lower vertebrates like Xenopus frog and Medaka fish possesses the binding affinity to methylated CpG dinucleotides in the same extent of those in mammals. Initially we addressed this as we previously identified a indispensable role of xl.Mbd4 with MLH1 for apoptotic induction in xDnmt1-depleted embryos (Ruzov et al., 2009). Partial depletion of xDnmt1 in Xenopus is sufficient to induce apoptosis without general hypo-methylation (Ruzov et al., 2009). However, If xl.Mbd4 could recognise mCpG dinucleotides was unknown. Thus it was interesting to investigate this point. Moreover, later I reasoned MBD4 might have a key role in Xenopus development, possibly in the context of active demethylation at thyroid hormone receptors and/or other hormone receptors. There are several piece of evidence supporting this point. While mice lacking *Mbd4* are viable, the morpholino knockdowns of *Mbd4* in Xenopus result in a lethal

phenotype in their early development, which fail to develop to tadpole stages (Ruzov et al., 2009). This is a potential precise phenotype to study the biological role of MBD4. A clue is that the thyroid hormone (T3) is essential to control the anuran amphibian metamorphosis including the development of tadpole-specific organs and the limb development (Matsuda et al., 2009). This involves gene regulation by the T3 receptor (TR). TRs form heterodimmers with 9-cis retinoic acid receptors (RXRs), the type II hormone receptors. One of the other heterodimmer-forming hormone receptors with RXRs is the nuclear vitamin D receptor, which is parathyroid (PTH) hormone-induced (Lee and Lee Kraus, 2001). These dimers bind to the response element (RE), most of which coincide with the promoters of target genes, negatively regulating their transcriptional repression (Kim et al., 2009). Recently a demethylation role of MBD4 at promoter of this specific vitamin D receptor has been reported (Kim et al., 2009). Coincidentally a study in zebrafish embryos has suggested zMbd4 could excise the G:T mismatches derived from deamination of 5mC, implicating its possible role in demethylation in this lower vertebrate organism (Rai et al., 2008). However, whether MBD4 of low vertebrates could efficiently recognise 5mC was unknown, especially there is a noticeable poor conservation of the primary sequences of its MBD domain. My results show the binding affinities of MBD4 in the lower Vertebrates of xenopus frog and Medaka fish are well conserved, suggesting that a possible role of Mbd4 in recognition of 5mC in the vertebrate genomes to act as a component of the machinery in, but not limited to the apoptosis signaling and the possible demethylation. Actually the role of MBD4 in binding methylation marks and inducing transcriptional repression might be underestimated, thus these results answered a prerequisite question for confirming its evolutionary binding affinity.

6.3.2. Maintenance of fidelity and regulation of stability of DNA methylation: a possible role of MBD4 and MeCP2

At the time of implantation of the early stage of mouse embryonic development, the initial pattern of DNA methylation are established by the DNMT3 enzymes (Popp et al., 2010, Sasaki and Matsui, 2008, Lees-Murdock and Walsh, 2008, Morgan, 2005, Reik, 2001). Once the DNA methylation patterns are established, they must be maintained and replicated with fidelity. DNMT1 prefers hemimethylated DNA over an unmethylated counterpart (Figure.6.6 A), and reproduces the methylation pattern according to the hemimethylated template after each round of replication (Jurkowska et al., 2010). In addition, UHRF1 has

been shown to bind specifically to hemimethylated DNA through its SRA domain (Hashimoto et al., 2008, Bostick et al., 2007, Arita et al., 2008, Avvakumov et al., 2008), and target DNMT1 to these specific sites (Sharif et al., 2007, Bostick et al., 2007) (Figure. 6.6 A). DNMT1 and its association with UHRF1 perform a central role in maintenance of the patterns of DNA methylation after the synthesis of the new DNA double strands. Additionally MBD4 may regulate their activity by direct association with them, and/or through regulation of their ubiquitination by its interactions with USP7 and UHRF1 (Figure. 6.6 A and Chapter 4). One of the factors affecting the fidelity of the methylation marks at CpG sites is the spontaneous deamination of methylated cytosines. The 5-methyl-cytosine is prone to the spontaneous deamination (Figure. 6.6 lower left diagram), the most common single nucleotide mutation, which removes ammonia from the 5-methylcytosine, and results in thymine (Figure. 6.6 lower left diagram, process on the right). In contrast, deamination of unmethylated cytosine generates uracil (Figure. 6.6 lower left diagram, process on the left). MBD4 is one of the only two mammalian thymine glycosylases, and one of four uracil glycosylases. MBD4 has the MBD domain that has a preferential binding affinity to G:T context (Figure. 6.6 C1&C2), and its glycosylase domain specifically recognise and excise the thymine mismatches (Hendrich et al., 1999). Where thymine mismatches occur at CpG dinucleotides, thymine glycosylases like MBD4 may repair the mismatches, with the help of component proteins of BER pathway, and mismatch repair pathway proteins such as MLH1 (Figure. 6.6 C1&C2). In addition, the cytosine bases restored may be unmethylated. Thus these unmethylated cytosines at CpG context need to be remethylated faithfully to maintain the fidelity of the methylation information. This may be performed by a precise machinery in which MBD4 may be involved through its associations with DNMT1 and/or DNMT3b (Figure. 6.6 C1&C2).

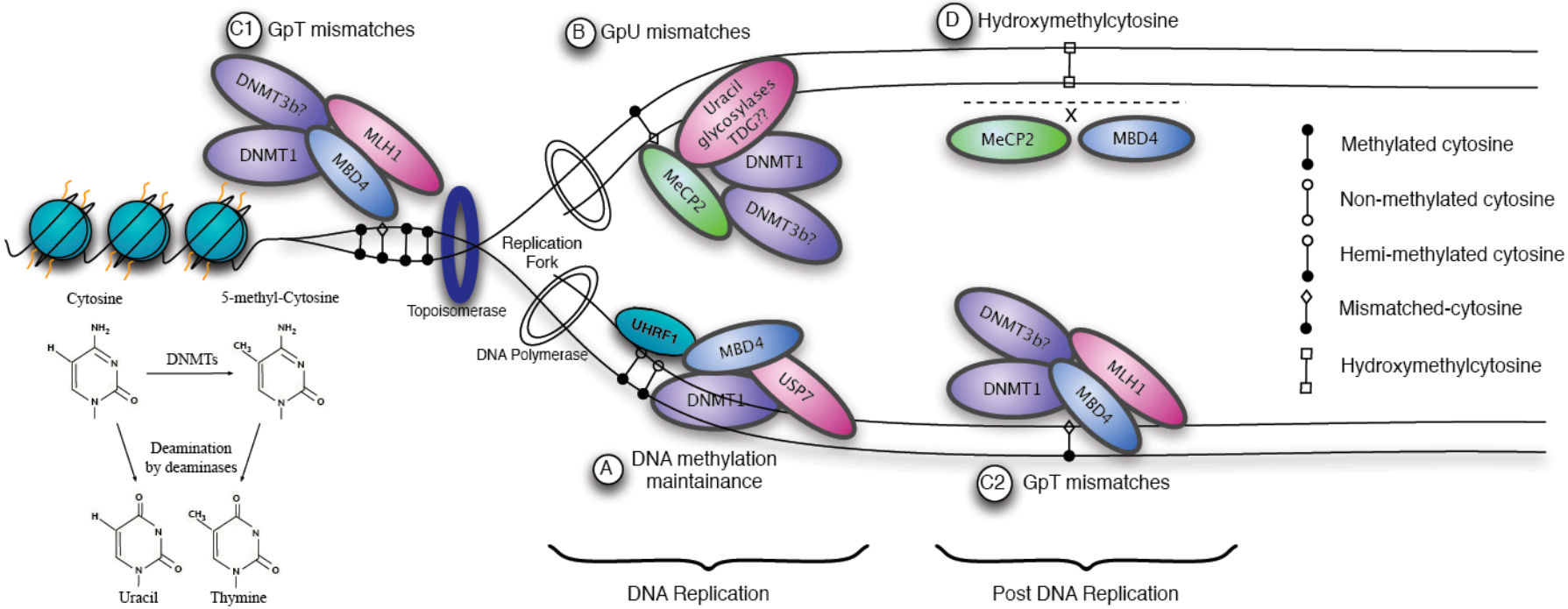

C1 GpT mismatches
DNMT3b?
DNMT1
MLH1
MBD4
B GpU mismatches
Uracil glycosylases TDG??
DNMT1
MeCP2
DNMT3b?
D Hydroxymethylcytosine
X
MeCP2
MBD4
Methylated cytosine
Non-methylated cytosine
Hemi-methylated cytosine
Mismatched-cytosine
Hydroxymethylcytosine
Cytosine
5-methyl-Cytosine
DNMTs
Deamination by deaminases
Uracil
Thymine
Topoisomerase
Replication Fork
DNA Polymerase
UHRF1
MBD4
USP7
DNMT1
A DNA methylation maintainance
DNMT3b?
DNMT1
MLH1
MBD4
C2 GpT mismatches
DNA Replication
Post DNA Replication

Figure 6.6. A working model for the role of MBD4 and MeCP2 in maintenance of fidelity and regulation of stability of DNA methylation. (A.) DNMT1 and UHRF1 can preferentially bind to hemi-methylated CpG dinucleotides. DNMT1 is present at the replication fork, and functions with the help of partner proteins including UHRF1 and PCNA (Proliferating cell nuclear antigen). MBD4 may regulate their activity by direct association with them, and/or through regulation of their ubiquitination by its interactions with USP7 and UHRF1. (B.) MeCP2 has binding preferences to a deamination mismatch product of uracil in the context of hemimethylated CpG dinucleotide. Other glycosylases (possibly TDG) might follow to excise and repair the uracil mismatches. The restored cytosine bases can be remethylated by their association proteins DNMT1 and/or DNMT3b. (C1&C2) The 5mC deamination product G:T mismatch can be preferentially recognized by MBD4, followed by excision and repair with the help of BER pathway and/or MMR pathway protein MLH1. MBD4 interacts directly with both DNMT1 (and/or DNMT3b) and MLH1 leading to recruitment of all of them at DNA damage sites. (D) Both MBD4 and MeCP2 are not capable of recognition of 5-hydroxymethylcytosine (5hmC) by themselves. However, this does not exclude they may involve in such processes where forming complexes with other capable proteins. Lower left diagram: components and processes of deamination of unmethylated and methylated cytosines.

I tested ol.Mbd4, xl.Mbd4 and xl.MeCP2 with hydroxymethylated (hmC) probes, and none of them showed binding specificity with hmC (Figure. 6.6 D), suggesting at least the MBD domains of these proteins could not distinguish this newly discovered epigenetic mark, while this does not exclude the possibility they might involve in recognition or excision of 5hmC in a complex with other possible 5hmC binding proteins. However, surprisingly I identified that the MBD domain of xl.MeCP2 has binding preferences to a deamination mismatch product of uracil in the context of hemimethylated CpG dinucleotide (Figure. 6.6 B). As far as I am aware, no published work has reported such binding specificity to deamination products like G:U by the MeCPs other than MBD4 and other glycosylase. The finding of this G:U mismatch affinity of MeCP2 with hemimethylated context has a very important implication that the MeCPs may involve in the efficient recognition of the deamination mismatches at CpG dinucleotides, to ensure the replication fidelity at the promoter-coincidence regions. There are a number of possible mechanisms proposed regarding how DNA glycosylases manage to efficiently search for and recognise the lesions including sliding-scan, tracking-diffusion-hopping, and forming a iron–sulfur cluster (Jacobs and Schär, 2011). However, if there is a preference repair to protect the most crucial sites like CpG dinucleotides and how this will work is unclear. In terms of hydrolytic deamination, MBD4 has demonstrated a binding preference to G:T mutation at CpG dinucleotides, and we have shown this is also true in Xenopus, suggesting a possible efficient repair mechanisms for this deamination product of 5mC (Figure. 6.6 C1&C2). This also in another way implicates a biological relevance of the special property of MBD4 in

possessing a MBD domain in addition to its glycosylase domain. Similarly, I have shown the G:U binding affinity of xl.MeCP2 is much stronger than its binding to the fully methylated probes, providing a preferential role of G:U recognision by MeCp2. Thus for the deamination product of unmethylated cytosine G:U at hemimethylated CpG sites, my results suggest a previously unknown machinery, in which G:U mismatches generated during DNA replication might be firstly recognised by MeCP2, and a uracil glycosylase such as TDG might follow, and induce BER repair for an unmethylated cytosine. Both MeCP2 and TDG have been reported to associate with DNA methyltransferases like DNMT1 or DNMT3b (Kimura, 2002, Boland and Christman, 2008), suggesting this speculation might be true. This also ensures the repaired cytosine at CpG sites can be remethylated subsequently to maintain the epigenetic stability in the genome (Figure. 6.6 B).

Chapter 7: General Discussion

My overall aim for this project was to identify new possible roles for MBD4, and the underlining mechanisms. I began by screening for MBD4 associated protein partners in two complementary high-throughput approaches of Y2H and AP/MS. Surprisingly a large number of protein candidates were identified from both screens. This led me to choose and validate a cohort of most interesting candidates by reciprocal co-IP and IF, followed by an identification of MBD4 domains responsible for their interactions. Almost all these identified MBD4 association proteins including UHRF1, PRMT5, KAP1, ACF1, and USP7 are known transcriptional co-repressors as histone modifiers and/or chromatin remodellers. This led me to investigate the role of MBD4 in transcriptional repression (TR) in the context of two independent *Mbd4*$^{-/-}$ MEF cell lines by performing microarrays and bioinformatic analysis to identify the possible MBD4 targets. Last but not least, I tested the binding specificity of MBD4 and its closest relative protein homologue MeCP2 against a variety of DNA probes containing unmethylated, methylated, hydroxymethylated, and mismatched cytosine bases at CpG dinucleotides. The main findings of this work are as follows:

- A large number of proteins (~380) potentially interact with MBD4, many of which are involved in the functional groups of signalling, transcriptional regulation, DNA repair and RNA processing.
- Majority of chosen candidates were validated by reciprocal co-IP of expressed partners and by immunofluorescence (IF) microscopy to determine their potential co-localisation in mouse and human cell lines, including UHRF1, PRMT5, KAP1, ACF1, and USP7.
- The intervening domain of MBD4 is a novel protein interaction region for tested candidates including UHRF1, PRMT5, and KAP1.
- The transcription regulatory components KAP1 and CFP1 not only interact with but also dramatically influence the stability of exogenously expressed MBD4 in human cells.
- A cohort of genes were identified that are significantly up- or down- regulated in both *Mbd4*$^{-/-}$ MEFs. This included mis-expression of insulin-like growth factor-binding proteins and two paternally imprinted genes Dio3 and H19. In response to this observation, I identified a potential interaction between MBD4 and estrogen receptor α (ERα) by co-IP and IF co-localisation.
- By pathway analysis, several developmental genes including *Sox9*, *Klf2* and *Klf4*, were prioritised as possible MBD4 targets. On this basis I propose a role for MBD4 in acquired diseases such as cancers and autoimmune diseases via transcriptional regulation.

- DNA binding specificity to a series of methylated and mismatched probes is conserved regardless of the poor sequence conservation of the MBD domain of Mbd4 between the species.

Based on above findings, I conclude that MBD4 is integrated in multiple pathways in the nucleus that includes DNA repair, chromatin remodelling, transcriptional regulation and genome stability. It is noteworthy that although MBD4 has a functionally conserved methyl-CpG binding domain, this is unlikely to be major methyl-CpG associated machinery. Because in both MEFs or normal and cancer cells from human and mouse, the methylation statuses of the promoter regions of the developmental genes that are potential MBD4 targets are usually low (Supplementary table 7.1). This may indicate indirect targeting by MBD4 and link with discussed protein interactors that are involved in transcriptional regulation. It is very likely that potential targeting at these promoters rely on the protein association complex including MBD4 and its co-repressors including KAP1, PRMT5, USP7 and UHRF1. The only exception is the two paternally imprinted genes Dio3 and H19, whose differentially methylated regions (DMRs) usually remain highly methylated.

Further work

In terms of immediate future work I think it will be interesting to check the endogenous MBD4 expression in human primary cancer tissues such as breast, colorectal and ovarian cancers. One of the limitations of my PhD work was the lack of a good MBD4 antibody that could check endogenous MBD4 expression. At the very end of this thesis writing, I noticed that there is a new human MBD4 antibody now available at Sigma. I have tried this new antibody, and observed both nuclear and cytoplasmic expression of MBD4 in human primary cancer tissues (Supplementary figure 7.1). In response to this observation, right before the writing of this general discussion chapter, I have performed immunofluorescent Aquanalysis to examine the association of expressions of MBD4 in line with TDG and SOX9, on tissue microarrays containing primary human cancer specimens of 182 breast cancers, 130 colorectal cancer, and 247 ovarian cancers (for breast cancers tissue microarray, we also have ERα and PgR expression data). This will be important to test the working model of the role of MBD4 in cancers regarding the possible mechanism of transcriptional repression and/or apoptosis signaling. These expression results need to be classed by their pathologic subtypes and/or tumour grades and/or prognosis and metastasis history. In addition, it is also very interesting to check my MBD4 association findings using this MBD4 antibody by endogenous co-IP in primary tissues.

It is very interesting to knockdown MBD4 in relevant cell lines such as MCF7 to further investigate transcriptional changes according to the *Mbd4*$^{-/-}$ MEFs microarray results. And as mentioned in above chapter, MBD4 rescue experiment is necessary to confirm the direct targets of MBD4. Importantly, repression assays for MBD4 need to be set up to ask if MBD4 and the identified co-repressive complexes indeed act on target genes in a manner of transcriptional repression. A Gal4 based vector system together with CREB and KAP1 vectors can be used as a system for transcriptional reporter assays. If this system could work for MBD4, it will be very interesting to see if the different combinations of MBD4 association proteins including KAP1, PRMT5, UHRF1 and ACF1 could differentially regulate the transcriptional level. In terms of the reporters that would be suitable for the assay, it is interesting to try the promoter regions of the developmental genes discussed including *SOX9*, *KLF2*, and *KLF4*, which have either NF-kB or cbp/p300 binding sites. This may accompany with EMSAs (electrophoretic mobility shift assay) using purified MBD4 with fluorescent or radioactive labeled probes of NF-kB and/or CREB binding sites within their promoters containing differential CpG dinucleotides. In addition, methylation status of the promoters of a number of potential MBD4 targets including the above-mentioned developmental genes and paternally imprinted genes need to be investigated by bisulfate sequencing and/or methyl-DIP.

In the longer term, it will be necessary to check the MBD4 binding sites in the genome by CHIP-seq in both cultured cells and primary tissues. The new MBD4 antibody may be suited for this task. Ideally it should be done in a more biologically relevant system such as certain cancer cell lines such as MCF7 and chondrocytes, and primary tissues such as breast cancers and/or synovial joints tissues from rheumatoid arthritis patients. Moreover, the effects of chronic stress on mice models that lacking MBD4 need to be re-evaluated, especially for the collagen antibody stress and estradiol stress, to test the hypothesis described in Figure. 5.10. In addition, it will be interesting to test DNA damage responses on *MBD4*$^{-/-}$ cells or MBD4 knockdown cells for the role of MBD4 in apoptotic responses, while keeping an eye on the status of p21 and *KLF4* expression. Last but not least, it will be interesting to knockout multiple MeCPs to see if this would cause more severe phenotype. A multiple knockout of MeCP2, Mbd2 and Kaiso in mice has been reported but the phenotype observed was very similar to the single knockout of MeCP2 (Martin Caballero et al., 2009). One of the possibilities would be other MeCPs existing in the system compensated their roles of transcriptional regulation, including MBD4. Thus more multiple knockouts of MeCPs including MBD4 would help understand the precise role of MeCPs in the genome in the future.

Appendix

Supplementary Table 3.1.1 Protein candidate list identified in Mbd4 full length Y2H screen.

Protein Name	Description (The proteins without the colony PCR hit number are the ones identified in plasmid rescues)	Colony PCR hits	Panther molecular function
mRNA transcription regulation			
E4f1	E4F transcription factor 1; May function as a transcriptional repressor. May also function as a ubiquitin ligase mediating ubiquitination of chromatin-associated TP53. Functions in cell survival and proliferation through control of the cell cycle. Functions in the p53 and pRB tumor suppressor pathways and regulates the cyclin CCNA2 transcription (782 aa)	10	Transcription factor Zinc finger transcription factor KRAB box transcription factor
Zscan22	Zinc finger and SCAN domain-containing protein 22 (Zinc finger protein 50) (Krueppel-related zinc finger protein 2) (Protein HKR2); May be involved in transcriptional regulation (491 aa)*(Homo sapiens)*	2	Transcription factor Zinc finger transcription factor KRAB box transcription factor
Zfp748	zinc finger protein 748 (888 aa)	1	Transcription factor Zinc finger transcription factor KRAB box transcription factor
Ring1	ring finger protein 1; Constitutes one of the E3 ubiquitin-protein ligases that mediate monoubiquitination of 'Lys-119' of histone H2A, thereby playing a central role in histone code and gene regulation. H2A 'Lys-119' ubiquitination gives a specific tag for epigenetic transcriptional repression and participates in X chromosome inactivation of female mammals. Essential component of the Polycomb group (PcG) multiprotein PRC1 complex, a complex required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development. PcG PRC1 complex act via c [...] (406 aa)	3	Transcription factor Zinc finger transcription factor
Pcgf6	polycomb group ring finger 6; Transcriptional repressor. May modulate the levels of histone H3K4Me3 by activating JARID1D histone demethylase (By similarity) (353 aa)	2	Transcription factor Zinc finger transcription factor
Rnf2	ring finger protein 2; E3 ubiquitin-protein ligase that mediates monoubiquitination of 'Lys-119' of histone H2A, thereby playing a central role in histone code and gene regulation. H2A 'Lys-119' ubiquitination gives a specific tag for epigenetic transcriptional repression and participates in X chromosome inactivation of female mammals. May be involved in the initiation of both imprinted and random X inactivation. Essential component of the Polycomb group (PcG) multiprotein PRC1 complex, a complex required to maintain the transcriptionally repressive state of many genes, including Hox g [...] (336 aa)	2	Transcription factor Zinc finger transcription factor
Zfp161	zinc finger protein 161; Transcriptional repressor of c-Myc and thymidine kinase promoters (448 aa)	1	Transcription factor Zinc finger transcription factor
Mysm1	myb-like, SWIRM and MPN domains 1 (819 aa)	1	Transcription factor Transcription cofactor Nucleic acid binding
Hoxb9	homeo box B9; Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis (250 aa)	1	Transcription factor Homeobox transcription factor Nucleic acid binding Other DNA-binding protein
Lhx1	LIM homeobox protein 1; Potential transcription factor. May play a role in early mesoderm formation and later in lateral mesoderm differentiation and neurogenesis (406 aa)	2	Transcription factor Homeobox transcription factor Zinc finger transcription factor Other zinc finger transcription factor Nucleic acid binding

Hes6	hairy and enhancer of split 6 (Drosophila); Does not bind DNA itself but suppresses both HES1- mediated N box-dependent transcriptional repression and binding of HES1 to E box sequences. Also suppresses HES1-mediated inhibition of the heterodimer formed by ASCL1/MASH1 and TCF3/E47, allowing ASCL1 and TCF3 to up-regulate transcription in its presence. Promotes cell differentation (224 aa)	1	Transcription factor Basic helix-loop-helix transcription factor Nucleic acid binding
Pja2	praja 2, RING-H2 motif containing; Has E2-dependent E3 ubiquitin-protein ligase activity (707 aa)	2	Transcription factor
mycn	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian) (462 aa)	3	Transcription factor Basic helix-loop-helix transcription factor Nucleic acid binding
Ewsr1	RNA-binding protein EWS (EWS oncogene) (Ewing sarcoma breakpoint region 1 protein); Might normally function as a repressor. EWS-fusion- proteins (EFPS) may play a role in the tumorigenic process. They may disturb gene expression by mimicking, or interfering with the normal function of CTD-POLII within the transcription initiation complex. They may also contribute to an aberrant activation of the fusion protein target genes (656 aa)*(Homo sapiens)*	1	Nucleic acid binding Other RNA-binding protein
Crebzf	CREB/ATF bZIP transcription factor; Strongly activates transcription when bound to HCFC1. Suppresses the expression of HSV proteins in cells infected with the virus in a HCFC1-dependent manner. Also suppresses the HCFC1-dependent transcriptional activation by CREB3 and reduces the amount of CREB3 in the cell. Able to down-regulate expression of some cellular genes in CREBZF-expressing cells (By similarity) (273 aa)		Transcription factor CREB transcription factor Nucleic acid binding
Lsm14b	LSM14 protein homolog B (Protein FAM61B) (385 aa)		Nucleic acid binding Other RNA-binding protein
Cir1	corepressor interacting with RBPJ, 1		Transcription factor Transcription cofactor
Ttc3	Tetratricopeptide repeat protein 3 (TPR repeat protein 3) (TPR repeat protein D) (RING finger protein 105) (2025 aa)Ttc3 tetratricopeptide repeat domain 3; contain RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc; defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)- H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C; probably involved in mediating protein-protein interactions; identified in a proteins with a wide range of functions such as viral replication, signal transduction, and development; has two variants, the C3HC4-type and a C3H2C3-type (RING-H2 finger), which have different cysteine/histidine pattern; a subset of RINGs are associated with B-Boxes (C-X2-H-X7-C-X7-C-X2-C-H-X2-H)	3	Molecular function unclassified
Cenpv	proline-rich polypeptide 6; Required for distribution of pericentromeric heterochromatin in interphase nuclei and for centromere formation and organization, chromosome alignment and cytokinesis (By similarity) (252 aa)	1	Molecular function unclassified
Uimc1	ubiquitin interaction motif containing 1; Acts as a transcriptional repressor. Inhibits the interaction of NR6A1 with the corepressor NCOR1 (By similarity) (729 aa)	1	Molecular function unclassified
Signaling, Transporter & Receptor			
Plcd1	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase delta 1 (EC 3.1.4.11) (Phosphoinositide phospholipase C) (PLC-delta-1) (Phospholipase C-delta-1) (PLC-III); The production of the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) is mediated by activated phosphatidylinositol-specific phospholipase C enzymes. Essential for trophoblast and placental development (777 aa)*(Homo sapiens)*	2	Hydrolase Lipase Phospholipase Select calcium binding protein
Fbn2	fibrillin 2; Fibrillins are structural components of 10-12 nm extracellular calcium-binding microfibrils, which occur either in association with elastin or in elastin-free bundles. Fibrillin-2- containing microfibrils regulate the early process of elastic fiber assembly (2909 aa)	1	Signaling molecule Cell adhesion molecule Extracellular matrix Extracellular matrix structural protein Extracellular matrix glycoprotein Select calcium binding protein Annexin
Hba-a1	hemoglobin alpha, adult chain 1 (Hba-a1)	2	Transfer/carrier protein Other transfer/carrier protein
Slc25a37	solute carrier family 25, member 37; Mitochondrial iron transporter that specifically mediates iron uptake in developing erythroid cells, thereby playing an essential role in heme biosynthesis. The iron delivered into the mitochondria,	2	Transfer/carrier protein Mitochondrial carrier protein

	presumably as Fe(2+), is then probably delivered to ferrochelatase to catalyze Fe(2+) incorporation into protoprophyrin IX to make heme (338 aa)		
G3bp2	Ras GTPase activating protein (SH3 domain) binding protein 2; Probable scaffold protein that may be involved in mRNA transport (Potential) (482 aa)	1	Signaling molecule Other signaling molecule Nucleic acid binding Other RNA-binding protein
Nck2	non-catalytic region of tyrosine kinase adaptor protein 2 (380 aa)	1	Signaling molecule Other signaling molecule Miscellaneous function Other miscellaneous function protein Transmembrane receptor regulatory/adaptor protein
Marcks	myristoylated alanine rich protein kinase C substrate; MARCKS is the most prominent cellular substrate for protein kinase C. This protein binds calmodulin, actin, and synapsin. MARCKS is a filamentous (F) actin cross-linking protein (309 aa)	2	Signaling molecule Other signaling molecule Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein Miscellaneous function Structural protein
Lgals2	lectin, galactose-binding, soluble 2; This protein binds beta-galactoside. Its physiological function is not yet known (130 aa)	1	Signaling molecule Other signaling molecule Cell adhesion molecule Other cell adhesion molecule
Rab11fip3	RAB11 family interacting protein 3 (class II) (907 aa)	1	Signaling molecule Other signaling molecule Select regulatory molecule Other enzyme regulator Select calcium binding protein Calmodulin related protein Miscellaneous function
Vps16	vacuolar protein sorting 16 (yeast); May play a role in vesicle-mediated protein trafficking to lysosomal compartments and in membrane docking/fusion reactions of late endosomes/lysosomes (By similarity) (839 aa)	6	Molecular function unclassified
Stx8	syntaxin 8; Vesicle trafficking protein that functions in the early secretory pathway, possibly by mediating retrograde transport from cis-Golgi membranes to the ER (By similarity) (236 aa)	1	Membrane traffic protein SNARE protein
Arf6	ADP-ribosylation factor 6; GTP-binding protein that functions as an allosteric activator of the cholera toxin catalytic subunit, an ADP- ribosyltransferase. Involved in protein trafficking; may modulate vesicle budding and uncoating within the Golgi apparatus (By similarity) (175 aa)	1	Select regulatory molecule G-protein Small GTPase
Lrrn1	Leucine-rich repeat neuronal protein 1 precursor (Neuronal leucine- rich repeat protein 1) (NLRR-1)*(Homo sapiens)*	1	Receptor Extracellular matrix
Kit	kit oncogene; This is the receptor for stem cell factor (mast cell growth factor). It has a tyrosine-protein kinase activity. Binding of the ligands leads to the autophosphorylation of KIT and its association with substrates such as phosphatidylinositol 3-kinase (Pi3K) (979 aa)	1	Receptor Protein kinase receptor Tyrosine protein kinase receptor Kinase Protein kinase
Loxl1	lysyl oxidase-like 1; Active on elastin and collagen substrates (By similarity) (607 aa)	2	Receptor Other receptor Oxidoreductase Oxidase
Nr1h3	nuclear receptor subfamily 1, group H, member 3; Orphan receptor. Interaction with RXR shifts RXR from its role as a silent DNA-binding partner to an active ligand- binding subunit in mediating retinoid responses through target genes defined by LXRES. LXRES are DR4-type response elements characterized by direct repeats of two similar hexanuclotide half- sites spaced by four nucleotides. Plays an important role in the regulation of cholesterol homeostasis (445 aa)	1	Receptor Nuclear hormone receptor Transcription factor Nucleic acid binding
Pth1r	parathyroid hormone receptor 1; This is a receptor for parathyroid hormone and for parathyroid hormone-related peptide. The activity of this receptor is mediated by G proteins which activate adenylyl cyclase and also a phosphatidylinositol-calcium second messenger system (591 aa)	2	Receptor G-protein coupled receptor

Rhoa	ras homolog gene family, member A; Regulates a signal transduction pathway linking plasma membrane receptors to the assembly of focal adhesions and actin stress fibers. May be an activator of PLCE1 (By similarity) (193 aa)	1	Select regulatory molecule G-protein Small GTPase
Ran	RAN, member RAS oncogene family (Ran)	1	Select regulatory molecule G-protein Small GTPase
Dcaf8	DDB1 and CUL4 associated factor 8 (Dcaf8)	1	Select regulatory molecule G-protein Miscellaneous function Other miscellaneous function protein
Gnb2	guanine nucleotide binding protein, beta 2; Guanine nucleotide-binding proteins (G proteins) are involved as a modulator or transducer in various transmembrane signaling systems. The beta and gamma chains are required for the GTPase activity, for replacement of GDP by GTP, and for G protein-effector interaction (340 aa)	1	Select regulatory molecule G-protein Large G-protein Hydrolase
Atl3	atlastin GTPase 3 (Atl3)	1	Select regulatory molecule G-protein Large G-protein
Atp1b2	ATPase, Na+/K+ transporting, beta 2 polypeptide; This is the non-catalytic component of the active enzyme, which catalyzes the hydrolysis of ATP coupled with the exchange of Na(+) and K(+) ions across the plasma membrane. The exact function of the beta-2 subunit is not known (290 aa)		Transporter Cation transporter
Atp1b3	ATPase, Na+/K+ transporting, beta 3 polypeptide; This is the non-catalytic component of the active enzyme, which catalyzes the hydrolysis of ATP coupled with the exchange of Na(+) and K(+) ions across the plasma membrane. The exact function of the beta-3 subunit is not known (278 aa)		Transporter Cation transporter
Exoc5	exocyst complex component 5; Component of the exocyst complex involved in the docking of exocystic vesicles with fusion sites on the plasma membrane (By similarity) (709 aa)		Membrane traffic protein Membrane traffic regulatory protein Other membrane traffic protein
Mapkapk5	MAP kinase-activated protein kinase 5 (EC 2.7.11.1) (MAPK-activated protein kinase 5) (MAPKAP kinase 5) (p38-regulated/activated protein kinase); Mediates stress-induced small heat shock protein 27 phosphorylation (461 aa)		Kinase Protein kinase Non-receptor serine/threonine protein kinase
Odz3	Teneurin-3 (Ten-3) (Tenascin-M3) (Ten-m3) (Protein Odd Oz/ten-m homolog 3); May function as a cellular signal transducer (2715 aa)		Receptor Other receptor Signaling molecule Other signaling molecule Membrane-bound signaling molecule
Hba-a2	hemoglobin alpha, adult chain 2; Involved in oxygen transport from the lung to the various peripheral tissues (142 aa)		Transfer/carrier protein Other transfer/carrier protein
Slc35b4	solute carrier family 35, member B4; Sugar transporter that specifically mediates the transport of UDP-xylose (UDP-Xyl) and UDP-N-acetylglucosamine (UDP-GlcNAc) from cytosol into Golgi (By similarity) (331 aa)		Transporter Carbohydrate transporter Other transporter
Tmed2	transmembrane emp24 domain trafficking protein 2; Could have a role in the budding of coatomer-coated and other species of coated vesicles. Could bind cargo molecules to collect them into budding vesicles (By similarity) (201 aa)		Membrane traffic protein Vesicle coat protein
Efcab7	Efcab7 EF-hand calcium binding domain 7; EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF hands. Ca2+ binding induces a conformational change in the EF-hand motif, leading to the activation or inactivation of target proteins. EF-hands tend to occur in pairs or higher copy numbers.	4	Molecular function unclassified
Gpr149	Probable G-protein coupled receptor 149 (G-protein coupled receptor PGR10); Orphan receptor (730 aa)	1	Molecular function unclassified
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH); Independent of its glycolytic activity it is also involved in membrane trafficking in the early secretory pathway (335 aa)	2	Molecular function unclassified
Hmmr	hyaluronan mediated motility receptor (RHAMM); Involved in cell motility. When hyaluronan binds to HMMR, the phosphorylation of a number of proteins, including the focal adhesion kinase occurs. May also be involved in cellular transformation and metastasis formation, and in regulating extracellular-regulated kinase (ERK) activity (794 aa)	2	Molecular function unclassified

Protein synthesis regulation, modification & Proteolysis			
Eif3a	eukaryotic translation initiation factor 3, subunit 10 (theta); Binds to the 40S ribosome and promotes the binding of methionyl-tRNAi and mRNA. May act as an adapter molecule between the translation initiation apparatus and the cytoskeleton structure in the cell (By similarity) (1344 aa)	15	Nucleic acid binding Translation factor Translation initiation factor
Eef1a1	eukaryotic translation elongation factor 1 alpha 1; This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis (By similarity) (468 aa)	5	Nucleic acid binding Translation factor Translation elongation factor
Cacybp	calcyclin binding protein; May be involved in calcium-dependent ubiquitination and subsequent proteosomal degradation of target proteins. Probably serves as a molecular bridge in ubiquitin E3 complexes. Participates in the ubiquitin-mediated degradation of beta-catenin (CTNNB1) (By similarity) (229 aa)	1	Ligase Ubiquitin-protein ligase
Prmt10	protein arginine methyltransferase 10 (putative) (Prmt10)	1	Transferase Methyltransferase
Sephs1	selenophosphate synthetase 1; Synthesizes selenophosphate from selenide and ATP (392 aa)	2	Transferase Other transferase
Tpp2	tripeptidyl peptidase II; Component of the proteolytic cascade acting downstream of the 26S proteasome in the ubiquitin-proteasome pathway. May be able to complement the 26S proteasome function to some extent under conditions in which the latter is inhibited (By similarity) (1262 aa)	3	Protease Serine protease
Psma5	proteasome (prosome, macropain) subunit, alpha type 5; The proteasome is a multicatalytic proteinase complex which is characterized by its ability to cleave peptides with Arg, Phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or slightly basic pH. The proteasome has an ATP-dependent proteolytic activity (By similarity) (241 aa)	2	Protease Other proteases
Bmp1	bone morphogenetic protein 1; Cleaves the C-terminal propeptides of procollagen I, II and III. Induces cartilage and bone formation. May participate in dorsoventral patterning during early development by cleaving chordin (CHRD) (By similarity) (991 aa)	5	Protease Metalloprotease
Usp7	Ubiquitin carboxyl-terminal hydrolase 7 (EC 3.1.2.15) (Ubiquitin thioesterase 7) (Ubiquitin-specific-processing protease 7) (Deubiquitinating enzyme 7) (Herpesvirus-associated ubiquitin-specific protease); Cleaves ubiquitin fusion protein substrates. Deubiquitinates TP53/p53 and MDM2 and strongly stabilizes TP53 even in the presence of excess MDM2, and also induces TP53-dependent cell growth repression and apoptosis (1110 aa)*(Homo sapiens)*	1	Protease Cysteine protease
Phf21b	PHD finger protein 21B (505 aa)	3	Ligase Ubiquitin-protein ligase
Anapc2	anaphase promoting complex subunit 2; Component of the anaphase promoting complex/cyclosome (APC/C), a cell cycle-regulated ubiquitin ligase that controls progression through mitosis and the G1 phase of the cell cycle (By similarity) (837 aa)	1	Ligase Ubiquitin-protein ligase
Anapc10	anaphase promoting complex subunit 10; Component of the anaphase promoting complex/cyclosome (APC/C), a cell cycle-regulated ubiquitin ligase that controls progression through mitosis and the G1 phase of the cell cycle (By similarity) (185 aa)	2	Select regulatory molecule Ligase Other ligase
Cul1	cullin 1; Core component of multiple cullin-RING-based SCF (SKP1- CUL1-F-box protein) E3 ubiquitin-protein ligase complexes, which mediate the ubiquitination of proteins involved in cell cycle progression, signal transduction and transcription. In the SCF complex, serves as a rigid scaffold that organizes the SKP1-F-box protein and RBX1 subunits. May contribute to catalysis through positioning of the substrate and the ubiquitin-conjugating enzyme. The E3 ubiquitin-protein ligase activity of the complex is dependent on the neddylation of the cullin subunit and is inhibited by the associ [...] (776 aa)	2	Molecular function unclassified
Skp1a	S-phase kinase-associated protein 1A; Essential component of the SCF (SKP1-CUL1-F-box protein) ubiquitin ligase complex, which mediates the ubiquitination of proteins involved in cell cycle progression, signal transduction and transcription. In the SCF complex, serves as an adapter that links the F-box protein to CUL1 (By similarity) (163 aa)	1	Ligase Ubiquitin-protein ligase
Nedd8	neural precursor cell expressed, developmentally down-regulated gene 8; Ubiquitin-like protein which plays an important role in cell cycle control and embryogenesis. Covalent attachment to its substrates requires prior activation by the E1 complex UBE1C- APPBP1 and linkage to the E2 enzyme UBE2M. Attachment of NEDD8 to cullins activates their associated E3 ubiquitin ligase	1	Nucleic acid binding Ribosomal protein

	activity, and thus promotes polyubiquitination and proteasomal degradation of cyclins and other regulatory proteins (75 aa)		
Ube2e1	ubiquitin-conjugating enzyme E2E 1, UBC4/5 homolog (yeast); Catalyzes the covalent attachment of ubiquitin to other proteins. Mediates the selective degradation of short-lived and abnormal proteins (193 aa)	2	Ligase Other ligase
Papss1	3'-phosphoadenosine 5'-phosphosulfate synthase 1; Bifunctional enzyme with both ATP sulfurylase and APS kinase activity, which mediates two steps in the sulfate activation pathway. The first step is the transfer of a sulfate group to ATP to yield adenosine 5'-phosphosulfate (APS), and the second step is the transfer of a phosphate group from ATP to APS yielding 3'-phosphoadenylylsulfate (PAPS- activated sulfate donor used by sulfotransferase). In mammals, PAPS is the sole source of sulfate; APS appears to be only an intermediate in the sulfate- activation pathway. Also involved in the [...] (627 aa)	1	Kinase Other kinase Transferase Nucleotidyltransferase
Cndp2	CNDP dipeptidase 2 (metallopeptidase M20 family) (475 aa)		Protease Metalloprotease
Pcbp2	poly(rC) binding protein 2; Single-stranded nucleic acid binding protein that binds preferentially to oligo dC (362 aa)		Select regulatory molecule
Pias1	protein inhibitor of activated STAT 1; Functions as an E3-type small ubiquitin-like modifier (SUMO) ligase, stabilizing the interaction between UBE2I and the substrate, and as a SUMO-tethering factor. Plays a crucial role as a transcriptional coregulation in various cellular pathways, including the STAT pathway, the p53 pathway and the steroid hormone signaling pathway. The effects of this transcriptional coregulation, transactivation or silencing, may vary depending upon the biological context (644 aa)		Ligase
Ptpn2	Tyrosine-protein phosphatase non-receptor type 2 (EC 3.1.3.48) (T-cell protein-tyrosine phosphatase) (TCPTP) (415 aa)		Phosphatase Protein phosphatase
Trim27	tripartite motif protein 27; Has a transcriptional repressor activity. Induces apoptosis by activating Jun N-terminal kinase and p38 kinase and also increases caspase-3-like activity independently of mitochondrial events. May function in male germ cell development. Has DNA-binding activity and preferentially bound to double- stranded DNA (By similarity) (521 aa)		Ligase Ubiquitin-protein ligase
Ube2m	ubiquitin-conjugating enzyme E2M (UBC12 homolog, yeast); Accepts the ubiquitin-like protein NEDD8 from the UBA3- NAE1 E1 complex and catalyzes its covalent attachment to other proteins. Involved in cell proliferation (By similarity) (183 aa)		Ligase Other ligase
Cstb	Cystatin-B (Stefin-B) (Liver thiol proteinase inhibitor) (CPI-B); This is an intracellular thiol proteinase inhibitor. Tightly binding reversible inhibitor of cathepsins L, H and B (98 aa)		Select regulatory molecule Protease inhibitor Cysteine protease inhibitor
Naa50	N(alpha)-acetyltransferase 50, NatE catalytic subunit		Transferase Acetyltransferase
Sugt1	Suppressor of G2 allele of SKP1 homolog (Sgt1) (Putative 40-6-3 protein); May play a role in ubiquitination and subsequent proteosomal degradation of target proteins (365 aa)	2	Molecular function unclassified
Zc3hc1	zinc finger, C3HC type 1; Essential component of an SCF-type E3 ligase complex, SCF(NIPA), a complex that controls mitotic entry by mediating ubiquitination and subsequent degradation of cyclin B1 (CCNB1). Its cell-cycle-dependent phosphorylation regulates the assembly of the SCF(NIPA) complex, restricting CCNB1 ubiquitination activity to interphase. Its inactivation results in nuclear accumulation of CCNB1 in interphase and premature mitotic entry (By similarity). Overexpression may be able to protect from apoptosis induced by IL-3 withdrawal (502 aa)	2	Molecular function unclassified
Ribosomal protein			
Rpl12	ribosomal protein (Rpl12)	32	Nucleic acid binding Ribosomal protein
Ubc	ubiquitin C; Protein modifier which can be covalently attached to target lysines either as a monomer or as a lysine-linked polymer. Attachment to proteins as a Lys-48-linked polymer usually leads to their degradation by proteasome. Attachment to proteins as a monomer or as an alternatively linked polymer does not lead to proteasomal degradation and may be required for numerous functions, including maintenance of chromatin structure, regulation of gene expression, stress response, ribosome biogenesis and DNA repair (By	28	Nucleic acid binding Ribosomal protein

	similarity) (886 aa)		
Rps20	ribosomal protein S20 (Rps20)	8	Nucleic acid binding Ribosomal protein
Ubb	ubiquitin B (Ubb)	3	Nucleic acid binding Ribosomal protein
Brix1	BRX1, biogenesis of ribosomes, homolog (S. cerevisiae) (Brix1)	1	Nucleic acid binding Ribosomal protein
Rpl5	ribosomal protein L5 (Rpl5)	1	Nucleic acid binding Ribosomal protein
Rpl7a	ribosomal protein L7A (Rpl7a)	1	Nucleic acid binding Ribosomal protein
Rps6	ribosomal protein S6; May play an important role in controlling cell growth and proliferation through the selective translation of particular classes of mRNA (249 aa)	1	Molecular function unclassified
	-		Pre-mRNA processing
Cpsf3	cleavage and polyadenylation specificity factor 3; Component of the cleavage and polyadenylation specificity factor (CPSF) complex that play a key role in pre-mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with poly(A) polymerase and other factors to bring about cleavage and poly(A) addition. Has endonuclease activity, and functions as mRNA 3'-end-processing endonuclease (By similarity) (684 aa)	1	Nucleic acid binding Nuclease Endoribonuclease mRNA processing factor mRNA polyadenylation factor
Sfrs2	splicing factor, arginine/serine-rich 2 (SC-35); Necessary for the splicing of pre-mRNA. It is required for formation of the earliest ATP-dependent splicing complex and interacts with spliceosomal components bound to both the 5'- and 3'-splice sites during spliceosome assembly. It also is required for ATP-dependent interactions of both U1 and U2 snRNPs with pre- mRNA (By similarity). Can bind to the myelin basic protein (MBP) gene MB3 regulatory region and increase transcription of the mbp promoter in cells derived from the CNS (277 aa)	116	Nucleic acid binding mRNA processing factor mRNA splicing factor
Sfrs5	splicing factor, arginine/serine-rich 5 (SRp40, HRS); May be required for progression through G1 and entry into S phase of cell growth. May play a regulatory role in pre- mRNA splicing. Autoregulates its own expression. Plays a role in constitutive splicing and can modulate the selection of alternative splice sites (By similarity) (269 aa)	24	Nucleic acid binding mRNA processing factor mRNA splicing factor
Sfrs7	splicing factor, arginine/serine-rich 7; Required for pre-mRNA splicing (By similarity) (266 aa)	3	Nucleic acid binding mRNA processing factor mRNA splicing factor
Sfrs15	splicing factor, arginine/serine-rich 15 (1199 aa)	7	Nucleic acid binding mRNA processing factor
Sfrs18	Splicing factor, arginine/serine-rich 130 (Serine-arginine-rich- splicing regulatory protein 130) (SRrp130) (SR-rich protein) (SR- related protein) (805 aa)	1	Molecular function unclassified
Fip1l1	FIP1_MOUSE Isoform 2 of Q9D824 - Mus musculus (Mouse); Component of the cleavage and polyadenylation specificity factor (CPSF) complex that plays a key role in pre- mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with poly(A) polymerase and other factors to bring about cleavage and poly(A) addition. FIP1L1 contributes to poly(A) site recognition and stimulates poly(A) addition. Binds to U-rich RNA sequence elements surrounding the poly(A) site. May act to tether poly(A) polymerase to the CPSF complex (By similarity) (545 aa)	17	Molecular function unclassified
Ints7	integrator complex subunit 7; Component of the Integrator complex, a complex involved in the small nuclear RNAs (snRNA) U1 and U2 transcription and in their 3'-box-dependent processing. The Integrator complex is associated with the C-terminal domain (CTD) of RNA polymerase II largest subunit (POLR2A) and is recruited to the U1 and U2 snRNAs genes (By similarity) (965 aa)	6	Molecular function unclassified
DNA binding, replication, repair & Chromatin binding			

Taf1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor; Largest component and core scaffold of the TFIID basal transcription factor complex. Contains novel N- and C-terminal Ser/Thr kinase domains which can autophosphorylate or transphosphorylate other transcription factors. Phosphorylates TP53 on 'Thr-55' which leads to MDM2-mediated degradation of TP53. Phosphorylates GTF2A1 and GTF2F1 on Ser residues. Possesses DNA- binding activity. Essential for progression of the G1 phase of the cell cycle (By similarity) (1891 aa)	1	Transcription factor Nucleic acid binding Nuclease Kinase Protein kinase Transferase Acetyltransferase Acyltransferase
Setd2	Histone-lysine N-methyltransferase SETD2 (EC 2.1.1.43) (SET domain-containing protein 2) (hSET2) (Huntingtin-interacting protein HYPB) (Huntingtin yeast partner B) (Huntingtin-interacting protein 1) (HIF- 1) (p231HBP); Histone methyltransferase that methylates 'Lys-36' of histone H3. H3 'Lys-36' methylation represents a specific tag for epigenetic transcriptional activation. Probably plays a role in chromatin structure modulation during elongation via its interaction with hyperphosphorylated POLR2A. Binds DNA at promoters. May also act as a transcription activator that binds to promot [...] (2061 aa)*(Homo sapiens)*	1	Nucleic acid binding Other DNA-binding protein Transferase Methyltransferase
G2e3	G2/M-phase specific E3 ubiquitin ligase (G2e3)	1	Nucleic acid binding Double-stranded DNA binding protein
Polr1d	DNA-directed RNA polymerase I subunit D (EC 2.7.7.6) (DNA-directed RNA polymerase I 16 kDa polypeptide) (RPA16); DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. Common core component of RNA polymerases I and III which synthesize ribosomal RNA precursors and small RNAs, such as 5S rRNA and tRNAs, respectively (133 aa)*(Homo sapiens)*	1	Nucleic acid binding DNA-directed RNA polymerase Transferase Nucleotidyltransferase
Top2b	DNA topoisomerase 2-beta (EC 5.99.1.3) (DNA topoisomerase II, beta isozyme); Control of topological states of DNA by transient breakage and subsequent rejoining of DNA strands. Topoisomerase II makes double-strand breaks (By similarity) (1626 aa)	2	Nucleic acid binding DNA topoisomerase Select regulatory molecule Isomerase
Pcna	proliferating cell nuclear antigen; This protein is an auxiliary protein of DNA polymerase delta and is involved in the control of eukaryotic DNA replication by increasing the polymerase's processibility during elongation of the leading strand (By similarity) (261 aa)	22	Nucleic acid binding DNA polymerase processivity factor
Sec61b	Sec61 beta subunit; Necessary for protein translocation in the endoplasmic reticulum (By similarity) (122 aa)	2	Nucleic acid binding DNA ligase Ligase
Hmgn2	Nonhistone chromosomal protein HMG-17 (High-mobility group nucleosome-binding domain-containing protein 2); Binds to the inner side of the nucleosomal DNA thus altering the interaction between the DNA and the histone octamer. May be involved in the process which maintains transcribable genes in an unique chromatin conformation (By similarity) (90 aa)*(Homo sapiens)*	1	Nucleic acid binding Chromatin/chromatin-binding protein
Sap18	Histone deacetylase complex subunit SAP18 (Sin3-associated polypeptide, 18 kDa) (Sin3-associated polypeptide p18) (2HOR0202); Component of the SIN3-repressing complex. Enhances the ability of SIN3-HDAC1-mediated transcriptional repression. When tethered to the promoter, it can direct the formation of a repressive complex to core histone proteins (153 aa)*(Homo sapiens)*	1	Nucleic acid binding Chromatin/chromatin-binding protein
Fus	RNA-binding protein FUS (Oncogene FUS) (Oncogene TLS) (Translocated in liposarcoma protein) (POMp75) (75 kDa DNA-pairing protein); Binds both single-stranded and double-stranded DNA and promotes ATP-independent annealing of complementary single- stranded DNAs and D-loop formation in superhelical double-stranded DNA. May play a role in maintenance of genomic integrity (526 aa)*(Homo sapiens)*	1	Transcription factor Other transcription factor Nucleic acid binding mRNA processing factor mRNA splicing factor Other DNA-binding protein
Pcbp1	poly(rC) binding protein 1; Single-stranded nucleic acid binding protein that binds preferentially to oligo dC (By similarity) (356 aa)	2	Select regulatory molecule
Ctbp1	C-terminal binding protein 1; Involved in controlling the equilibrium between tubular and stacked structures in the Golgi complex (By similarity). Co-repressor targeting diverse transcription regulators such as GLIS2. Has dehydrogenase activity (441 aa)		Transcription factor Transcription cofactor Oxidoreductase Dehydrogenase
Dek	DEK oncogene (DNA binding); May have a function in the nucleus (380 aa)		Nucleic acid binding DNA ligase Ligase

Rpa1	Replication protein A 70 kDa DNA-binding subunit (RP-A) (RF-A) (Replication factor-A protein 1) (Single-stranded DNA-binding protein) (p70); It participates in a very early step in initiation. RP-A is a single-stranded DNA-binding protein. Absolutely required for simian virus 40 DNA replication in vitro (616 aa)		Nucleic acid binding Single-stranded DNA-binding protein
Setd1a	Histone-lysine N-methyltransferase, H3 lysine-4 specific SET1 (EC 2.1.1.43) (Set1/Ash2 histone methyltransferase complex subunit SET1) (SET domain-containing protein 1A); Histone methyltransferase that specifically methylates 'Lys-4' of histone H3, when part of the SET1 histone methyltransferase (HMT) complex, but not if the neighboring 'Lys- 9' residue is already methylated. H3 'Lys-4' methylation represents a specific tag for epigenetic transcriptional activation. The non-overalpping localization with SETD1B suggests that SETD1A and SETD1B make non-redundant contributions to the epig [...] (1709 aa)		Nucleic acid binding Other DNA-binding protein Transferase Methyltransferase
Zfand5	Mus musculus zinc finger, AN1-type domain 5 (Zfand5), mRNA		Nucleic acid binding Miscellaneous function Other miscellaneous function protein
Zfp280b	suppressor of hairy wing homolog 2 (Drosophila) (534 aa)		Transcription factor Zinc finger transcription factor
Ss18	SSXT protein (Synovial sarcoma, translocated to X chromosome) (SYT protein) (418 aa)		Nucleic acid binding Chromatin/chromatin-binding protein
Xrcc1	X-ray repair complementing defective repair in Chinese hamster cells 1; Corrects defective DNA strand-break repair and sister chromatid exchange following treatment with ionizing radiation and alkylating agents (631 aa)		Nucleic acid binding
Aplf	aprataxin and PNKP like factor; Involved in single-strand and double-strand DNA break repair (By similarity) (499 aa)	1	Molecular function unclassified
Men1	multiple endocrine neoplasia 1; May be involved in DNA repair. Component of MLL- containing complexes (named MLL, ASCOM, MLL2/MLL3 or MLL3/MLL4 complex)- at least composed ASH2L, RBBP5, DPY30, WDR5, one or several histone methyltransferases (MLL, MLL2, MLL3 and/or MLL4), and the facultative components MEN1, HCFC1, HCFC2, NCOA6, UTX, PAXIP1/PTIP and C16orf53/PA1 (616 aa)		Molecular function unclassified
Cell skeleton, adhesion & Chaperone			
Emilin1	elastin microfibril interfacer 1; May be responsible for anchoring smooth muscle cells to elastic fibers, and may be involved not only in the formation of the elastic fiber, but also in the processes that regulate vessel assembly. Has cell adhesive capacity (By similarity). May have a function in placenta formation and initial organogenesis and a later role in interstitial connective tissue (1017 aa)	1	Extracellular matrix Extracellular matrix glycoprotein
Fcho2	FCH domain only 2; May play a role in membrane remodeling by imposing and stabilizing particular membrane curvatures (By similarity) (781 aa)	1	Cytoskeletal protein Actin binding cytoskeletal protein Other actin family cytoskeletal protein Membrane traffic protein Membrane traffic regulatory protein
Flna	filamin, alpha; Promotes orthogonal branching of actin filaments and links actin filaments to membrane glycoproteins. Anchors various transmembrane proteins to the actin cytoskeleton and serves as a scaffold for a wide range of cytoplasmic signaling proteins (By similarity). Interaction with FLNA may allow neuroblast migration from the ventricular zone into the cortical plate. Tethers cell surface-localized furin, modulates its rate of internalization and directs its intracellular trafficking (2647 aa)	6	Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein
Flnb	Filamin-B (FLN-B) (Beta-filamin) (Actin-binding-like protein) (Thyroid autoantigen) (Truncated actin-binding protein) (Truncated ABP) (ABP- 280 homolog) (ABP-278) (Filamin 3) (Filamin homolog 1) (Fh1); Connects cell membrane constituents to the actin cytoskeleton. May promote orthogonal branching of actin filaments and links actin filaments to membrane glycoproteins. Anchors various transmembrane proteins to the actin cytoskeleton. Interaction with FLNA may allow neuroblast migration from the ventricular zone into the cortical plate. Various interactions and localizations of isoforms a [...] (2602 aa)*(Homo sapiens)*	2	Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein
Fhod1	formin homology 2 domain containing 1; Required for the assembly of F-actin structures, such as stress fibers. Depends on the Rho-ROCK cascade for its activity. Contributes to the coordination of microtubules with actin fibers and	1	Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein

	plays a role in cell elongation (By similarity) (1197 aa)		
Whrn	whirlin; Necessary for elongation and maintenance of inner and outer hair cell stereocilia in the organ of Corti in the inner ear (918 aa)	1	Cytoskeletal protein
Gypa	Glycophorin A precursor (PAS-2) (Sialoglycoprotein alpha) (MN sialoglycoprotein) (CD235a antigen); Glycophorin A is the major intrinsic membrane protein of the erythrocyte. The N-terminal glycosylated segment, which lies outside the erythrocyte membrane, has MN blood group receptors and also binds influenza virus (150 aa)*(Homo sapiens)*	1	Molecular function unclassified
Cdh11	cadherin 11; Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types (796 aa)	4	Cell adhesion molecule Cadherin
Pcdhgb6	protocadherin gamma subfamily C, 3 (944 aa)	1	Cell adhesion molecule Cadherin
Pcdhgb7	protocadherin gamma subfamily B, 7 (Pcdhgb7)	2	Cell adhesion molecule Cadherin
Ahsa1	AHA1, activator of heat shock protein ATPase homolog 1 (yeast); Cochaperone that stimulates HSP90 ATPase activity. May affect a step in the endoplasmic reticulum to Golgi trafficking (By similarity) (338 aa)	1	Chaperone Other chaperones Select regulatory molecule Other enzyme regulator Other enzyme activator
Fkbp3	FK506 binding protein 3; FK506- and rapamycin-binding proteins (FKBPs) constitute a family of receptors for the two immunosuppressants which inhibit T-cell proliferation by arresting two distinct cytoplasmic signal transmission pathways. PPIases accelerate the folding of proteins (By similarity) (224 aa)	4	Chaperone Other chaperones Isomerase Other isomerase
Dnajb1	DnaJ (Hsp40) homolog, subfamily B, member 1; Interacts with HSP70 and can stimulate its ATPase activity. Stimulates the association between HSC70 and HIP (By similarity) (340 aa)	2	Chaperone Other chaperones
Dnaja3	DnaJ (Hsp40) homolog, subfamily A, member 3; Modulates apoptotic signal transduction or effector structures within the mitochondrial matrix. Affect cytochrome C release from the mitochondria and caspase 3 activation, but not caspase 8 activation. Isoform 1 increases apoptosis triggered by both TNF and the DNA-damaging agent mytomycin C; in sharp contrast, isoform 2 suppresses apoptosis. Can modulate IFN-gamma- mediated transcriptional activity (By similarity) (480 aa)	1	Chaperone Other chaperones
Dnajb4	DnaJ homolog subfamily B member 4 (Heat shock 40 kDa protein 1 homolog) (Heat shock protein 40 homolog) (HSP40 homolog) (337 aa)*(Homo sapiens)*	1	Chaperone Other chaperones
Dnajc18	DnaJ (Hsp40) homolog, subfamily C, member 18 (Dnajc18)	1	Chaperone Other chaperones
Grp94	Endoplasmin precursor (Heat shock protein 90 kDa beta member 1) (94 kDa glucose-regulated protein) (GRP94) (gp96 homolog) (Tumor rejection antigen 1); Molecular chaperone that functions in the processing and transport of secreted proteins. Has ATPase activity (By similarity) (803 aa)*(Homo sapiens)*	1	Chaperone Hsp 90 family chaperone
Hspa5	heat shock 70kD protein 5 (glucose-regulated protein); Probably plays a role in facilitating the assembly of multimeric protein complexes inside the ER, AND MASS SPECTROMETRY (655 aa)	1	Chaperone Hsp 70 family chaperone
Hspa8	Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8); Chaperone. Isoform 2 may function as an endogenous inhibitory regulator of HSC70 by competing the co-chaperones (646 aa)*(Homo sapiens)*	2	Chaperone Hsp 70 family chaperone
Hspa12b	heat shock protein 12B	1	Chaperone Hsp 70 family chaperone
Fnbp4	formin binding protein 4 (1077 aa)		Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein
Mtap1b	microtubule-associated protein 1 B; The function of brain MAPS is essentially unknown. Phosphorylated MAP1B may play a role in the cytoskeletal changes that accompany neurite extension. Possibly MAP1B Binds to at least two tubulin subunits in the polymer, and this bridging of subunits might be involved in nucleating microtubule polymerization and in stabilizing microtubules (2464 aa)		Cytoskeletal protein Microtubule family cytoskeletal protein Non-motor microtubule binding protein
Fermt3	Fermt3 fermitin family homolog 3 (Drosophila) [Mus musculus]; Probably involved in cell adhesion (By similarity) (665 aa)	1	Molecular function unclassified
Hspa14	heat shock 70kDa protein 14 isoform 1 (509 aa)	2	Molecular function unclassified

Others			
Qars	glutaminyl-tRNA synthetase (775 aa)	1	Synthase and synthetase Synthetase Aminoacyl-tRNA synthetase
Paics	Multifunctional protein ADE2 [Includes- Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6) (SAICAR synthetase); Phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) (AIR carboxylase) (AIRC)] (451 aa)*(Homo sapiens)*	8	Synthase and synthetase Synthase Ligase
Pop7	processing of precursor 7, ribonuclease P family, (S. cerevisiae) (Pop7)	1	Nucleic acid binding Nuclease Endoribonuclease Hydrolase
Cmpk1	cytidylate kinase; Catalyzes specific phosphoryl transfer from ATP to UMP and CMP (By similarity) (227 aa)	6	Kinase Nucleotide kinase
Pgm2	Phosphoglucomutase-2 (EC 5.4.2.2) (Glucose phosphomutase 2) (PGM 2); This enzyme participates in both the breakdown and synthesis of glucose (By similarity) (612 aa)*(Homo sapiens)*	7	Isomerase Mutase
Eno1	179aa	6	Carbohydrate metabolism Glycolysis
Bpgm	2,3-bisphosphoglycerate mutase; Plays a major role in regulating hemoglobin oxygen affinity as a consequence of controlling 2,3-BPG concentration. Can also catalyze the reaction of EC 5.4.2.1 (mutase) and EC 3.1.3.13 (phosphatase), but with a reduced activity (259 aa)	5	Isomerase Mutase
Lactb	Serine beta-lactamase-like protein LACTB, mitochondrial precursor (EC 3.4.-.-) (547 aa)*(Homo sapiens)*	1	Hydrolase Other hydrolase Protease Serine protease
Crmp1	collapsin response mediator protein 1; Necessary for signaling by class 3 semaphorins and subsequent remodeling of the cytoskeleton. Plays a role in axon guidance, invasive growth and cell migration (574 aa)*(Mus musculus)*	1	Hydrolase Other hydrolase
Obsl1	Obscurin-like protein 1 precursor	1	Hydrolase Other hydrolase
Txnl1	thioredoxin-like 1 (289 aa)	1	Oxidoreductase Other oxidoreductase
R3hdm1	R3H domain 1 (binds single-stranded nucleic acids) (R3hdm1)	2	Molecular function unclassified
Sms	spermine synthase; Required for normal viability, growth and fertility (366 aa)	1	Synthase and synthetase Synthase
Npc2	Niemann Pick type C2; May be involved in the regulation of the lipid composition of sperm membranes during the maturation in the epididymis (By similarity) (149 aa)	4	Molecular function unclassified
Aldh9a1	4-trimethylaminobutyraldehyde dehydrogenase (EC 1.2.1.47) (TMABADH) (Aldehyde dehydrogenase 9A1) (EC 1.2.1.3) (Aldehyde dehydrogenase E3 isozyme) (Gamma-aminobutyraldehyde dehydrogenase) (EC 1.2.1.19) (R-aminobutyraldehyde dehydrogenase); Converts gamma-trimethylaminobutyraldehyde into gamma- butyrobetaine. Catalyzes the irreversible oxidation of a broad range of aldehydes to the corresponding acids in an NAD-dependent reaction (518 aa)		Oxidoreductase Dehydrogenase
Cyb5r4	cytochrome b5 reductase 4; NADH-cytochrome b5 reductase involved in endolasmic reticulum stress response pathway. Plays a critical role in protecting pancreatic beta-cells against oxidant stress, possibly by protecting the cell from excess buildup of reactive oxygen species (ROS) (497 aa)		Oxidoreductase Reductase
Fech	Ferrochelatase, mitochondrial precursor (EC 4.99.1.1) (Protoheme ferro-lyase) (Heme synthetase); Catalyzes the ferrous insertion into protoporphyrin IX (By similarity) (429 aa)		Lyase Other lyase
RNASEH2A	ribonuclease H2, large subunit; Catalytic subunit of RNase HII, an endonuclease that specifically degrades the RNA of RNA-DNA hybrids. Participates in DNA replication, possibly by mediating the removal of lagging-strand Okazaki fragment RNA primers during DNA replication. Mediates the excision of single ribonucleotides from DNA-RNA duplexes (By similarity) (301 aa)		Nucleic acid binding Nuclease Endoribonuclease Hydrolase
Exosc5	exosome component 5; Component of the exosome 3'->5' exoribonuclease complex, a complex that degrades inherently unstable mRNAs containing AU-rich elements (AREs) within their 3'-untranslated regions. Required for the 3'-		Nucleic acid binding Nuclease Exoribonuclease

	processing of the 7S pre-RNA to the mature 5.8S rRNA. Has a 3'-5' exonuclease activity (By similarity) (235 aa)		Transferase Nucleotidyltransferase Hydrolase
Pgm1	phosphoglucomutase 2; This enzyme participates in both the breakdown and synthesis of glucose (By similarity) (562 aa)		Isomerase Mutase
Mybpc1	Myosin-binding protein C, slow-type (Slow MyBP-C) (C-protein, skeletal muscle slow isoform); Thick filament-associated protein located in the crossbridge region of vertebrate striated muscle a bands. In vitro it binds MHC, F-actin and native thin filaments, and modifies the activity of actin-activated myosin ATPase. It may modulate muscle contraction or may play a more structural role (1171 aa)	3	Molecular function unclassified
Poldip2	Polymerase delta-interacting protein 2 (38 kDa DNA polymerase delta interaction protein) (p38)	2	Molecular function unclassified
Golga1	golgi autoantigen, golgin subfamily a, 1; Probably involved in maintaining Golgi structure. May play a role in acrosome formation and spermatogenesis (758 aa)	1	Molecular function unclassified
Ngly1	N-glycanase 1; Specifically deglycosylates the denatured form of N- linked glycoproteins in the cytoplasm and assists their proteasome-mediated degradation. Cleaves the beta-aspartyl- glucosamine (GlcNAc) of the glycan and the amide side chain of Asn, converting Asn to Asp. Prefers proteins containing high- mannose over those bearing complex type oligosaccharides. Can recognize misfolded proteins in the endoplasmic reticulun that are exported in the cytosol to be destroyed and deglycosylate them, while it has no activity toward native proteins. Deglycosylation is prerequisite for subse [...] (651 aa)	1	
Tmem55a	transmembrane protein 55A; Catalyzes the hydrolysis of the 4-position phosphate of phosphatidylinositol 4,5-bisphosphate. Does not hydrolyze phosphatidylinositol 3,4,5-trisphosphate, phosphatidylinositol 3,4-bisphosphate, inositol 3,5-bisphosphate, inositol 3,4- bisphosphate, phosphatidylinositol 5-monophosphate, phosphatidylinositol 4-monophosphate and phosphatidylinositol 3- monophosphate (By similarity) (257 aa)		Molecular function unclassified
Peli2	pellino 2; Scaffold protein which probably links Toll-like receptors (TLRs) to basic cellular processes via its interaction with the complex containing IRAK kinases and TRAF6. Can activate the MAP (mitogen activated protein) kinase pathway leading to activation of ELK1. Not required for NF-kappa-B activation (419 aa)		Molecular function unclassified
Cenpc1	centromere protein C1; Component of the CENPA-NAC (nucleosome-associated) complex, a complex that plays a central role in assembly of kinetochore proteins, mitotic progression and chromosome segregation. The CENPA-NAC complex recruits the CENPA-CAD (nucleosome distal) complex and may be involved in incorporation of newly synthesized CENPA into centromeres (By similarity) (904 aa)		Molecular function unclassified
	-		Molecular function unclassified
Fam96b	Fam96b family with sequence similarity 96, member B [Mus musculus]	10	Molecular function unclassified
Fam128b	Fam128b family with sequence similarity 128, member B [Mus musculus]	3	Molecular function unclassified
Cep192	CEP192 centrosomal protein 192kDa [Homo sapiens]	2	Molecular function unclassified
N4bp2l2	N4bp2l2 NEDD4 binding protein 2-like 2 [Mus musculus]; contain NK; Nucleoside/nucleotide kinase (NK) is a protein superfamily consisting of multiple families of enzymes that share structural similarity and are functionally related to the catalysis of the reversible phosphate group transfer from nucleoside...	2	Molecular function unclassified
Phf13	PHD finger protein 13	2	Molecular function unclassified
Ehbp1l1	EH domain binding protein 1-like 1 (1718 aa); contain Calponin homology domain; actin-binding domain which may be present as a single copy or in tandem repeats (which increases binding affinity). The CH domain is found in cytoskeletal and signal transduction proteins, including actin-binding proteins like spectrin, alpha-actinin, dystrophin, utrophin, and fimbrin, proteins essential for regulation of cell shape (cortexillins), and signaling proteins (Vav).	1	Molecular function unclassified
Fndc3c1	Fndc3c1 fibronectin type III domain containing 3C1 [Mus musculus]; contain Fibronectin type 3 domain; One of three types of internal repeats found in the plasma protein fibronectin. Its tenth fibronectin type III repeat contains an RGD cell recognition sequence in a flexible loop between 2 strands. Approximately 2% of all animal proteins contain the FN3 repeat; including extracellular and intracellular proteins, membrane spanning cytokine receptors, growth hormone	1	Molecular function unclassified

	receptors, tyrosine phosphatase receptors, and adhesion molecules. FN3-like domains are also found in bacterial glycosyl hydrolases.		
Lrrc61	Leucine-rich repeat-containing protein 61 (259 aa)	1	Molecular function unclassified
Igsf3	immunoglobulin superfamily, member 3 isoform 2 (1214 aa)	5	Molecular function unclassified
Gm15450	Gm15450 predicted gene 15450 [Mus musculus]; contain Ribosomal_S23; S12-like family, 40S ribosomal protein S23 subfamily; S23 is located at the interface of the large and small ribosomal subunits of eukaryotes, adjacent to the decoding center. It interacts with domain III of the eukaryotic elongation factor 2 (eEF2),...	1	Molecular function unclassified
Zc4h2	Mus musculus zinc finger, C4H2 domain containing (Zc4h2), mRNA >gi|51261758|gb|BC079862.1| Mus musculus zinc finger, C4H2 domain containing, mRNA (cDNA clone MGC:105164 IMAGE:6807437), complete cds		Molecular function unclassified
Rsbn1	rosbin, round spermatid basic protein 1 (795 aa)		Molecular function unclassified
No Known Names			
AB114630	1 gene (Pcdha6: protocadherin alpha 6)	1	
AC007995	2 genes (Arhgap26: Rho GTPase activating protein 26; Nr3c1: nuclear receptor subfamily 3, group C, member 1.)	1	
AC083890	3 genes (Cutl1: cut-like 1 (Drosophila); Sh2b2: SH2B adaptor protein 2; Rpl7a: ribosomal protein L7a)	2	
AC098731	1 gene (Rbms3: RNA binding motif, single stranded interacting protein)	3	
AC112979	1 gene (EGF-like domain-containing protein C3orf50 homolog isoform b precursor)	1	
AC121523	1 gene (Dcc: deleted in colorectal carcinoma)	2	
AC122343	2 genes (Dkk3: dickkopf homolog 3 (*Xenopus laevis*); Usp47: ubiquitin specific peptidase 47)	2	
AC122903	1 gene (Ptprc: protein tyrosine phosphatase, receptor type, C)	1	
AC123954	7 genes (Bicd2: bicaudal D homolog 2 (Drosophila); Dcun1d1: DCUN1D1 DCN1, defective in cullin neddylation 1, domain containing 1 (S. cerevisiae); Ecm2: extracellular matrix protein 2, female organ and adipocyte specific; Ippk: inositol 1,3,4,5,6-pentakisphosphate 2-kinase; Cenpp: centromere protein P; Aspn: asporin; Dcun1d2: DCN1, defective in cullin neddylation 1, domain containing 2 (S. cerevisiae))	1	
AC125466	2 genes (Crb1: crumbs homolog 1 (Drosophila); Rbms3: RNA binding motif, single stranded interacting protein)	1	
AC130843	1 gene (Akt3: thymoma viral proto-oncogene 3)	1	
AC132370	1 gene (Dach1: dachshund 1 (Drosophila))	1	
AC151264	5 genes (Ube2c: ubiquitin-conjugating enzyme E2C; Heatr5b: HEAT repeat containing 5B; Strn: striatin, calmodulin binding protein; Mtap7: microtubule-associated protein 7; Ccdc75: coiled-coil domain containing 75)	1	
AC152164	no novel gene	1	
AC153508	3 genes (Appl2: adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 2; Aldh1l2: aldehyde dehydrogenase 1 family, member L2; WASH complex subunit 7)	2	
AC154436	4 genes (Rft1: RFT1 homolog (S. cerevisiae) ; Sfrs10: splicing factor, arginine/serine-rich 10; Tmem110: transmembrane protein 110; Sfmbt1: Scm-like with four mbt domains 1)	1	
AC154438	2 gene (Opa1: optic atrophy 1 homolog (human); Crxos1: Crx opposite strand transcript 1)	1	
AC157557	2 genes (Rps4x: ribosomal protein S4, X-linked; Nox4: NADPH oxidase 4)	1	
AC161812	1 gene (Acvr1b: activin A receptor, type 1B)	1	
AC164550	2 genes (Pnpla8: patatin-like phospholipase domain containing 8; Nrcam: neuron-glia-CAM-related cell adhesion molecule)	1	
AC164576	1 gene (Slc35f1: solute carrier family 35, member F1)	1	
AC165266	1 gene (Txndc10: thioredoxin domain containing 10)	1	
AC166834	4 genes (Dkk3: dickkopf homolog 3 (*Xenopus laevis*); Usp47: ubiquitin specific peptidase 47; Tra2a: transformer 2 alpha homolog (Drosophila); Mical2: microtubule associated monoxygenase, calponin and LIM domain containing 2)	3	

AC167464	4 genes (Fam134a: family with sequence similarity 134, member A; Atg9a: Mus musculus autophagy-related 9A (yeast); Abcb6: ATP-binding cassette, sub-family B (MDR/TAP), member 6; Zfand2b: zinc finger, AN1 type domain 2B)	1
AF220294	10 genes (Hn1l: hematological and neurological expressed 1-like; Igfals: insulin-like growth factor binding protein, acid labile subunit; Mapk8ip3: mitogen-activated protein kinase 8 interacting protein 3; Spsb3: splA/ryanodine receptor domain and SOCS box containing 3; Nubp2: nucleotide binding protein 2; Mrps34: mitochondrial ribosomal protein S34; Nme3: expressed in non-metastatic cells 3; Hagh: hydroxyacyl glutathione hydrolase; Ift140: intraflagellar transport 140 homolog (Chlamydomonas); HN1-like protein mRNA)	1
AF285162	1 gene (Ubc: ubiquitin C)	1
AF483216	1 gene (H13: histocompatibility 13)	1
AJ132592	1 gene (ZNF281: zinc finger protein 281)	1
AL589651	Olfr1361; Olfr1359; Olfr11; Olfr1362; Olfr1360; Hist1h1b; Hist1h3h; Hist1h3i; Hist1h2bg; Hist1h2be; Hist1h3g; Hist1h3d; Hist1h2an; Hist1h2ak; Hist1h2ai; Hist1h3a; Hist1h3b; Hist1h3c; Hist1h3e; Hist1h2ae; Hist1h2ag; Hist1h3f; Hist1h2ad; Hist1h2bl; Hist1h2bn; Hist1h2bm; Hist1h2bp; Hist1h2ah; Hist1h2ao; Hist1h2ac; Hist1h2af; Hist1h2ab; Hist1h2bc;	1
AL596384	9 genes (Nfe2l1: nuclear factor, erythroid derived 2,-like 1; Pnpo: pyridoxine 5'-phosphate oxidase; Sp2: Sp2 transcription factor; Atad4: ATPase family, AAA domain containing 4; Cbx1: chromobox homolog 1 (Drosophila HP1 beta); Rps2: ribosomal protein S2; Uba52: ubiquitin A-52 residue ribosomal protein fusion product 1; Copz2: coatomer protein complex, subunit zeta 2; Cdk5rap3: CDK5 regulatory subunit associated protein 3)	1
AL671671	5 genes (Zswim5: zinc finger, SWIM domain containing 5; Hectd3: HECT domain containing 3; Urod: uroporphyrinogen decarboxylase; Gtf2b: general transcription factor IIB; Zswim4: zinc finger, SWIM domain containing 4)	1
AL845162	12 genes (H13: histocompatibility 13; Mcts2: malignant T cell amplified sequence 2; Rem1: rad and gem related GTP binding protein 1; Mcts1: malignant T cell amplified sequence 1; Defb29: defensin beta 29; Tmem167: transmembrane protein 167; Defb19: defensin beta 19; Defb21: defensin beta 21; Defb36: defensin beta 36; Defb45: defensin beta 45; Defb25: defensin beta 25; Defb28: defensin beta 28)	1
BC017141	hypothetical protein LOC75964	1
BC032970	no novel gene	3
BC059832	1 gene (Acvr1b: activin A receptor, type 1B)	1
BC086760	1 gene (Cdh11: cadherin 11)	1
NM_020616	proline-rich protein BCA3	2
NR_015554	no novel gene	1
X82564	1 gene (Lars2:leucyl-tRNA synthetase, mitochondrial)	1
1500001M20Rik	1 gene (MMP37-like protein, mitochondrial precursor)	
2510012J08Rik	1 gene (hypothetical protein LOC70312)	
AC113050	1 gene (Rgs7:regulator of G protein signaling 7;	
AC158964	2 genes (Scyl3: SCY1-like 3 (S. cerevisiae); Kifap3: kinesin-associated protein 3)	
AC160932	3 genes (Acp1: acid phosphatase 1, soluble; Sh3yl1: Sh3 domain YSC-like 1; Zfp120: zinc finger protein 120.)	
AC161177	1 gene (Dcc: deleted in colorectal carcinoma)	
AC099643	2 genes (Rps4x: ribosomal protein S4, X-linked; Nox4: NADPH oxidase 4.)	
AC133198	1 gene (Zdhhc14:zinc finger, DHHC domain containing 14)	
D930014E17Rik	1 Gene (proline-rich protein BCA3)	

	Signaling, Transporter & Receptor	34
	Cell skeleton, adhesion & Chaperone	24
	Protein synthesis regulation, modification & Proteolysis	27
	mRNA transcription regulation	20
	DNA binding, replication, repair & Chromatin binding	21
	Ribosomal protein	8
	Pre-mRNA processing	8
	Others	27
	Molecular function unclassified	13
	No known Names	48
	Total	230

Supplementary Table 3.1.2 Protein candidate list identified in MBD domain Y2H screen.

Protein Name	Description (The proteins without the colony PCR hit number are the ones identified in plasmid rescues)	Colony PCR hits	overlap with full length screen?	Panther molecular function
mRNA transcription regulation				
Ash2l	ash2 (absent, small, or homeotic)-like (Drosophila); Component of the Set1/Ash2 histone methyltransferase (HMT) complex, a complex that specifically methylates 'Lys-4' of histone H3, but not if the neighboring 'Lys-9' residue is already methylated (By similarity). May function as a transcriptional regulator. May play a role in hematopoiesis (631 aa)	1	n	Transcription factor Zinc finger transcription factor Other zinc finger transcription factor
Ring1	ring finger protein 1; Constitutes one of the E3 ubiquitin-protein ligases that mediate monoubiquitination of 'Lys-119' of histone H2A, thereby playing a central role in histone code and gene regulation. H2A 'Lys-119' ubiquitination gives a specific tag for epigenetic transcriptional repression and participates in X chromosome inactivation of female mammals. Essential component of the Polycomb group (PcG) multiprotein PRC1 complex, a complex required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development. PcG PRC1 complex act via c [...] (406 aa)	4	y	Transcription factor Zinc finger transcription factor
Mycn	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian) (462 aa)	2	y	Transcription factor Basic helix-loop-helix transcription factor Nucleic acid binding
Pja2	praja 2, RING-H2 motif containing; Has E2-dependent E3 ubiquitin-protein ligase activity (707 aa)	1	y	Transcription factor
Protein synthesis regulation, modification & Proteolysis				
Eif3a	eukaryotic translation initiation factor 3, subunit 10 (theta); Binds to the 40S ribosome and promotes the binding of methionyl-tRNAi and mRNA. May act as an adapter molecule between the translation initiation apparatus and the cytoskeleton structure in the cell (By similarity) (1344 aa)	13	y	Nucleic acid binding Translation factor Translation initiation factor
Cacybp	calcyclin binding protein; May be involved in calcium-dependent ubiquitination and subsequent proteosomal degradation of target proteins. Probably serves as a molecular bridge in ubiquitin E3 complexes. Participates in the ubiquitin-mediated degradation of beta-catenin (CTNNB1) (By similarity) (229 aa)	2	y	Ligase Ubiquitin-protein ligase
Cell skeleton, adhesion & Chaperone				

Emilin1	elastin microfibril interfacer 1; May be responsible for anchoring smooth muscle cells to elastic fibers, and may be involved not only in the formation of the elastic fiber, but also in the processes that regulate vessel assembly. Has cell adhesive capacity (By similarity). May have a function in placenta formation and initial organogenesis and a later role in interstitial connective tissue (1017 aa)	1	y	Extracellular matrix Extracellular matrix glycoprotein
Tyro3	Tyrosine-protein kinase receptor TYRO3 precursor (EC 2.7.10.1) (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase SKY) (Tyrosine- protein kinase DTK) (Protein-tyrosine kinase byk); May be involved in cell adhesion processes, particularly in the central nervous system. In case of filovirus infection, seems to function as a cell entry factor (890 aa)(Homo sapiens)	2	n	Receptor Protein kinase receptor Tyrosine protein kinase receptor Kinase Protein kinase
Tbcb	tubulin folding cofactor B; Binds to alpha-tubulin folding intermediates after their interaction with cytosolic chaperonin in the pathway leading from newly synthesized tubulin to properly folded heterodimer (By similarity). Involved in regulation of tubulin heterodimer dissociation. May function as a negative regulator of axonal growth (244 aa)	1	n	Chaperone Other chaperones Cytoskeletal protein Microtubule family cytoskeletal protein Non-motor microtubule binding protein
Signaling, Transporter & Receptor				
Marcks	myristoylated alanine rich protein kinase C substrate; MARCKS is the most prominent cellular substrate for protein kinase C. This protein binds calmodulin, actin, and synapsin. MARCKS is a filamentous (F) actin cross-linking protein (309 aa)	3	y	Signaling molecule Other signaling molecule Cytoskeletal protein Actin binding cytoskeletal protein Non-motor actin binding protein Miscellaneous function Structural protein
Frem1	FRAS1-related extracellular matrix protein 1 precursor (Protein QBRICK); Extracellular matrix protein that plays a role in epidermal differentiation and is required for epidermal adhesion during embryonic development (By similarity) (2181 aa)	1	n	Receptor Other receptor
Nr1h3	nuclear receptor subfamily 1, group H, member 3; Orphan receptor. Interaction with RXR shifts RXR from its role as a silent DNA-binding partner to an active ligand- binding subunit in mediating retinoid responses through target genes defined by LXRES. LXRES are DR4-type response elements characterized by direct repeats of two similar hexanuclotide half- sites spaced by four nucleotides. Plays an important role in the regulation of cholesterol homeostasis (445 aa)	1	y	Receptor Nuclear hormone receptor Transcription factor Nucleic acid binding
Arcn1	archain 1; The coatomer is a cytosolic protein complex that binds to dilysine motifs and reversibly associates with Golgi non- clathrin-coated vesicles, which further mediate biosynthetic protein transport from the ER, via the Golgi up to the trans Golgi network. Coatomer complex is required for budding from Golgi membranes, and is essential for the retrograde Golgi-to-ER transport of dilysine-tagged proteins. In mammals, the coatomer can only be recruited by membranes associated to ADP-ribosylation factors (ARFs), which are small GTP-binding proteins; the complex also influences the G [...] (510 aa)	1	n	Membrane traffic protein Vesicle coat protein
Efcab7	Efcab7 EF-hand calcium binding domain 7; EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF hands. Ca2+ binding induces a conformational change in the EF-hand motif, leading to the activation or inactivation of target proteins. EF-hands tend to occur in pairs or higher copy numbers.	3	y	Molecular function unclassified
Ribosomal protein				
Rps20	ribosomal protein S20 (Rps20)	5	y	Nucleic acid binding Ribosomal protein
Ubc	ubiquitin C; Protein modifier which can be covalently attached to target lysines either as a monomer or as a lysine-linked polymer. Attachment to proteins as a Lys-48-linked polymer usually leads to their degradation by proteasome. Attachment to proteins as a monomer or as an alternatively linked polymer does not lead to proteasomal degradation and may be required for numerous functions, including maintenance of chromatin structure, regulation of gene expression, stress response,	3	y	Nucleic acid binding Ribosomal protein

	ribosome biogenesis and DNA repair (By similarity) (886 aa)			
Pre-mRNA processing				
Sfrs2	splicing factor, arginine/serine-rich 2 (SC-35); Necessary for the splicing of pre-mRNA. It is required for formation of the earliest ATP-dependent splicing complex and interacts with spliceosomal components bound to both the 5'- and 3'-splice sites during spliceosome assembly. It also is required for ATP-dependent interactions of both U1 and U2 snRNPs with pre- mRNA (By similarity). Can bind to the myelin basic protein (MBP) gene MB3 regulatory region and increase transcription of the mbp promoter in cells derived from the CNS (277 aa)	41	y	Nucleic acid binding mRNA processing factor mRNA splicing factor
Sfrs5	splicing factor, arginine/serine-rich 5 (SRp40, HRS) (Sfrs5)	2	y	Nucleic acid binding mRNA processing factor mRNA splicing factor
Sfrs7	splicing factor, arginine/serine-rich 7 (Sfrs7)	1	y	Nucleic acid binding mRNA processing factor mRNA splicing factor
Fip1l1	FIP1_MOUSE Isoform 2 of Q9D824 - Mus musculus (Mouse); Component of the cleavage and polyadenylation specificity factor (CPSF) complex that plays a key role in pre- mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with poly(A) polymerase and other factors to bring about cleavage and poly(A) addition. FIP1L1 contributes to poly(A) site recognition and stimulates poly(A) addition. Binds to U-rich RNA sequence elements surrounding the poly(A) site. May act to tether poly(A) polymerase to the CPSF complex (By similarity) (545 aa)	2	y	Molecular function unclassified
DNA binding, replication, repair & Chromatin binding				
Pcna	proliferating cell nuclear antigen; This protein is an auxiliary protein of DNA polymerase delta and is involved in the control of eukaryotic DNA replication by increasing the polymerase's processibility during elongation of the leading strand (By similarity) (261 aa)	2	y	Nucleic acid binding DNA polymerase processivity factor
Hmgn2	Nonhistone chromosomal protein HMG-17 (High-mobility group nucleosome-binding domain-containing protein 2); Binds to the inner side of the nucleosomal DNA thus altering the interaction between the DNA and the histone octamer. May be involved in the process which maintains transcribable genes in an unique chromatin conformation (By similarity) (90 aa)(Homo sapiens)	5	y	Nucleic acid binding Chromatin/chromatin-binding protein
Pcbp1	poly(rC) binding protein 1; Single-stranded nucleic acid binding protein that binds preferentially to oligo dC (By similarity) (356 aa)	2	y	Select regulatory molecule
Others				
Acbd3	acyl-Coenzyme A binding domain containing 3; Involved in the maintenance of Golgi structure by interacting with giantin, affecting protein transport between the endoplasmic reticulum and Golgi (By similarity). Involved in hormone-induced steroid biosynthesis in testicular Leydig cells (525 aa)	1	n	Molecular function unclassified
Cdca2	cell division cycle associated 2; Regulator of chromosome structure during mitosis required for condensin-depleted chromosomes to retain their compact architecture through anaphase. Acts by mediating the recruitment of phopsphatase PP1-gamma subunit (PPP1CC) to chromatin at anaphase and into the following interphase. At anaphase onset, its association with chromatin targets a pool of PPP1CC to dephosphorylate substrates (By similarity) (823 aa)	1	n	Molecular function unclassified
Molecular function unclassified				
Fam96b	family with sequence similarity 96, member B	1	y	Molecular function unclassified
				No Known Names
9130401M01Rik	hypothetical protein LOC75758	2	n	Molecular function unclassified
AC098731	1 gene (Rbms3: RNA binding motif, single stranded interacting protein)	1	y	
AL92868	1 gene (Ttn: titin)	1	n	

1		
	Signaling, Transporter & Receptor	5
	Cell skeleton, adhesion & Chaperone	3
	Protein synthesis regulation, modification & Proteolysis	2
	mRNA transcription regulation	4
	DNA binding, replication, repair & Chromatin binding	3
	Ribosomal protein	2
	Pre-mRNA processing	4
	Others	2
	Molecular function unclassified	1
	No known Names	3
	Total	29

Supplementary Table 3.2 Protein candidate list identified in AP/MS screen. IP lane reference: (2A-2F: candidates from two-step TAP IP; EP2: endogenous IP using mouse Mbd4 antibody from Epigenetek; FIP1-FIP5: Candidates from one step IP using FLAG tag; SC1: endogenous IP using mouse Mbd4 antibody from Santa Cruz)

Protein Name	IP lane	GI ID	Description	Mass (da)	MudPIT Score	Queries matched
TRIM28	**FIP2**	**gi|1699027 gi|31544959**	in Y2H is TRIM27. nuclear corepressor KAP-1 [Homo sapiens] TRIM28 protein [Homo sapiens]	80621	221	20
USP9X	**2A**	**gi|1666075**	ubiquitin hydrolase [Homo sapiens]ubiquitin specific peptidase 9, X-linked [Homo sapiens]	292860	209	11
CDKN2A	**FIP5**	**gi|4502709**	cell division cycle 2 isoform 1 [Homo sapiens]	34131	205	14
Top2	**FIP1**	**gi|37231**	In control as well. DNA topoisomerase II [Homo sapiens]	183548	191	23
SAE1	**FIP2**	**gi|35830**	ubiquitin activating enzyme E1 [Homo sapiens]	118799	181	8
ZMAT3	**FIP2**	**gi|21626466**	matrin 3 [Homo sapiens]	95078	178	19
SRRT	**FIP2**	**gi|46812675**	serrate RNA effector molecule homolog (Arabidopsis) [Homo sapiens] arsenate resistance protein ARS2 [Homo sapiens]	90650	163	14
DARS2	**FIP3**	**gi|40789249**	aspartyl-tRNA synthetase 2, mitochondrial precursor [Homo sapiens]	74086	162	13
MBD3	**FIP5**	**gi|4505119**	methyl-CpG binding domain protein 3 [Homo sapiens]	33051	161	10
ELAVL1	**FIP5**	**gi|119589356**	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R), isoform CRA_b [Homo sapiens]	50732	155	10
SMCHD1	**FIP1**	**gi|148839305**	structural maintenance of chromosomes flexible hinge domain containing 1 [Homo sapiens]	227942	154	12
THOC2	**FIP1**	**gi|125656165**	THO complex 2 [Homo sapiens]	184541	152	10
HDAC2	**FIP3**	**gi|1667394 gi|116284376**	histone deacetylase 2 [Homo sapiens]	66294	150	11
ACAT1	**FIP4**	**gi|499158**	mitochondrial acetoacetyl-CoA thiolase [Homo sapiens]	45537	140	8
SMC2	**FIP1**	**gi|3851584**	chromosome-associated protein-E [Homo sapiens]	136266	136	13
SLC25A6	**FIP5**	**gi|15928608**	Slc25a37 and Slc35b4 in Y2H. Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6 [Homo sapiens]	33133	130	14
CANX	**FIP3**	**gi|180631**	calnexin [Homo sapiens]	41096	129	5
V_region	**EP2**	**gi|619430**	This CDS feature is included to show the translation of the corresponding V_region. Presently translation qualifiers on V_region features are illegal [Homo sapiens] in SC2 as well	13370	127	6
MCM6	**FIP2**	**gi|7427519**	minichromosome maintenance complex component 6 [Homo sapiens]	93801	123	5
SET	**2D**	**gi|338039**	SET translocation (myeloid leukemia-associated) isoform 1 [Homo sapiens]	32115	118	8
SNRNP200	**FIP1**	**gi|20521660 gi|40217847**	activating signal cointegrator 1 complex subunit 3-like 1 [Homo sapiens]	246006	117	17
USP48	**FIP1**	**gi|52630449**	ubiquitin specific protease 48 isoform a [Homo sapiens]	121066	117	7

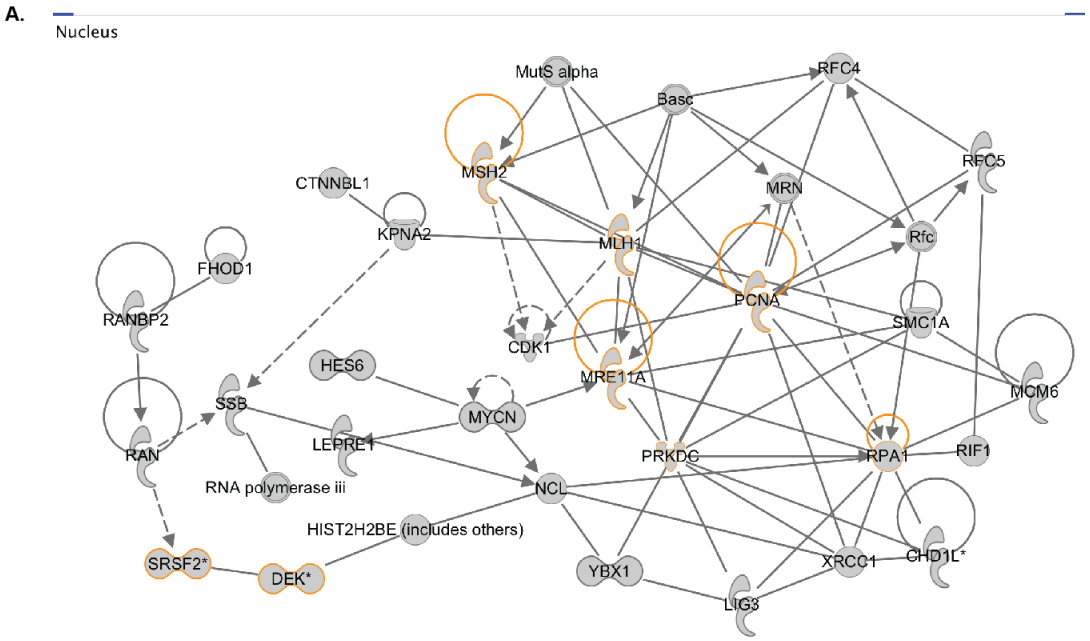
SLC25A5	**FIP5**	**gi|119610275**	Slc25a37 and Slc35b4 in Y2H. solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5, isoform CRA_b [Homo sapiens]	27843	116	17
GTF2B	**FIP5**	**gi|338043**	tat-associated protein (Transcription factor IIB serves as a bridge between IID, the factor which initially recognizes the promoter sequence, and RNA polymerase II.)	31364	112	4
USP9Y	**2A**	**gi|2181867**	DFFRY [Homo sapiens] ubiquitin specific protease 9 , Y-linked [Homo sapiens]	294468	107	6
VDAC3	**FIP5**	**gi|25188179**	voltage-dependent anion channel 3 isoform b [Homo sapiens]	30981	105	4
PRPF6	**FIP2**	**gi|4103604 gi|40807485**	putative mitochondrial outer membrane protein import receptor [Homo sapiens] PRP6 pre-mRNA processing factor 6 homolog [Homo sapiens]	107656	103	7
EIF3A	**FIP1**	**gi|116283747**	In control as well. EIF3A protein [Homo sapiens]	75019	101	11
RANBP2	**FIP1**	**gi|857368 gi|1009337**	RanBP2 (Ran-binding protein 2) [Homo sapiens]	362337	100	18
NCL	**2B**	**gi|189306**	nucleolin [Homo sapiens]	76355	100	6
LRRC15	**FIP5**	**gi|288541297**	in y2h is lrrc6. leucine rich repeat containing 15 isoform b [Homo sapiens]	65238	99	15
HADHA	**FIP3**	**gi|862457**	mitochondrial trifunctional protein, alpha subunit precursor [Homo sapiens] enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase alpha-subunit of trifunctional protein [Homo sapiens]	83688	98	8
RARS	**FIP3**	**gi|1217668**	arginyl-tRNA synthetase [Homo sapiens]	75614	96	7
SLC25A4	**FIP5**	**gi|339920**	Slc25a37 and Slc35b4 in Y2H. ADP/ADT translocator protein [Homo sapiens]solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4 [Homo sapiens]	33355	94	13
SFRS15	**FIP1**	**gi|4102967 gi|40789229**	pre-mRNA splicing SR protein rA4 [Homo sapiens] (splicing factor, arginine/serine-rich 15 isoform 1)	126132	91	7
DNM2	**FIP2**	**gi|56549123**	dynamin 2 isoform 2 [Homo sapiens]	98190	90	13
SF3A1	**FIP2**	**gi|5032087**	splicing factor 3b in control. splicing factor 3a, subunit 1, 120kDa isoform 1 [Homo sapiens]	88888	90	9
SRP68	**FIP3**	**gi|6690741**	signal recognition particle 68 [Homo sapiens]	70712	87	3
VDAC2	**FIP5**	**gi|119574954**	voltage-dependent anion channel 2, isoform CRA_a [Homo sapiens]	35029	85	7
CHD1L	**FIP2**	**gi|50418184**	CHD4 in control. chromodomain helicase DNA binding protein 1-like [Homo sapiens]	77878	85	6
SYMPK	**FIP1**	**gi|124028529**	symplekin [Homo sapiens]	141915	84	13
CYB5R3	**FIP5**	**gi|4503327**	cytochrome b5 reductase 3 isoform 1 [Homo sapiens]	34441	84	3
MTHFD1	**FIP2**	**gi|115206 gi|14602585**	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase [Homo sapiens]	102152	83	9
CSTF3	**FIP2**	**gi|4557495**	cleavage stimulation factor subunit 3 isoform 1 [Homo sapiens]	83325	83	6
BAZ1A	**FIP1**	**gi|6683494**	bromodomain adjacent to zinc finger domain 1A [Homo sapiens]	192610	79	14
EXOC2	**FIP2**	**gi|7023436**	Exoc5 in Y2H list. EXOC2 protein [Homo sapiens] Sec5 protein [Homo sapiens]	92115	78	1
NUMA1	**FIP1**	**gi|35121**	NuMA protein [Homo sapiens]	237523	77	13
DNMT1	**FIP1**	**gi|4503351**	In control as well. DNA (cytosine-5-)-methyltransferase 1 isoform b [Homo sapiens]	185388	77	11
MSH2	**FIP2**	**gi|4557761**	mutS homolog 2 [Homo sapiens] Chain A, Human Mutsalpha (Msh2MSH6) BOUND TO A G T MISPAIR, WITH Adp Bound To Msh2 Only	105418	77	10
IARS	**FIP1**	**gi|440799**	isoleucyl-tRNA synthetase [Homo sapiens]	145700	76	10
SMC4	**FIP1**	**gi|6807671 gi|50658063**	SMC4 structural maintenance of chromosomes 4-like 1 [Homo sapiens]	147775	75	8
TFRC	**FIP2**	**gi|37433 gi|189458817**	transferrin receptor [Homo sapiens]	85274	75	5
MOGS	**FIP2**	**gi|2344810**	a-glucosidase I [Homo sapiens]	91955	75	3
HADH	**FIP5**	**gi|5107725**	Chain A, Biochemical Characterization And Structure Determination Of Human Heart Short Chain L-3-Hydroxyacyl Coa Dehydrogenase Provide Insight Into Catalytic Mechanism hydroxyacyl-CoA dehydrogenase [Homo sapiens]	31848	74	9
AIMP2	**FIP5**	**gi|1215669**	aminoacyl tRNA synthetase complex-interacting multifunctional protein 2 [Homo sapiens]	35668	73	3
GTF3C5	**FIP3**	**gi|5281316 gi|170763508**	general transcription factor IIIC, polypeptide 5 isoform 1 [Homo sapiens]	60787	72	6
CHD3	**FIP1**	**gi|2645433**	CHD4 in control. CHD3 [Homo sapiens]chromodomain helicase DNA binding protein 3	222036	71	14

SMC1	**FIP1**	**gi|158257274 gi|999380**	mitosis-specific chromosome segregation protein SMC1 homolog [Homo sapiens]	143743	71	12
PDS5	**FIP1**	**gi|21951802 gi|155030216**	PDS5, regulator of cohesion maintenance, homolog A isoform 1 [Homo sapiens]	152274	71	11
RIF1	**FIP1**	**gi|55418562**	RAP1 interacting factor homolog (yeast) [Homo sapiens]	273726	71	6
ESYT1	**FIP1**	**gi|3882215 gi|3882215**	extended synaptotagmin-like protein 1 [Homo sapiens]	123293	71	3
HEATR1	**FIP1**	**gi|55959427**	HEAT repeat containing 1 [Homo sapiens]	235294	70	7
C4A	**EP2**	**gi|179674**	complement component C4A [Homo sapiens]	194337	70	2
COBRA1	**FIP3**	**gi|10434169 gi|20070260**	cofactor of BRCA1 [Homo sapiens]	66283	70	1
PTPLAD1	**FIP5**	**gi|6735452**	B-ind1 protein [Homo sapiens] protein tyrosine phosphatase-like A domain containing 1 [Homo sapiens]	43800	69	5
RFC4	**FIP5**	**gi|4506491**	replication factor C 4 [Homo sapiens]	40170	68	7
GTF3C3	**FIP2**	**gi|6808030 gi|28175202**	General transcription factor IIIC, polypeptide 3, 102kDa [Homo sapiens]	101979	67	5
ANXA2	**FIP5**	**gi|4757756**	annexin A2 isoform 2 [Homo sapiens]	38808	66	1
USP7	**FIP1**	**gi|1545952**	herpesvirus associated ubiquitin-specific protease (HAUSP) [Homo sapiens] double hits **gi|34851150** ubiquitin-specific protease 7 isoform [Homo sapiens]	129274	65	10
p67	**FIP2**	**gi|603225**	p67 [Homo sapiens]	65401	65	2
MTHFD2	**FIP5**	**gi|35071**	methylenetetrahydrofolate dehydrogenase 2 precursor variant [Homo sapiens]	37440	64	3
NR2F2	**FIP4**	**gi|30140**	nuclear receptor subfamily 2, group F, member 1 [Homo sapiens]	47095	63	1
TBL3	**FIP2**	**gi|19913369**	transducin beta-like 3 [Homo sapiens]	90347	62	5
RFC5	**FIP4**	**gi|6677723**	replication factor C 5 isoform 1 [Homo sapiens]	38757	60	1
ATP5C1	**FIP5**	**gi|4885079**	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma subunit isoform H (heart) precursor [Homo sapiens]	32917	59	4
GTF3C1	**FIP1**	**gi|50978466 gi|101943240**	general transcription factor IIIC, polypeptide 1, alpha 220kDa [Homo sapiens]	241062	58	12
LBP1d	**FIP3**	**gi|476099**	LBP-1d=transcription factor binding to initiation site of HIV-1 alternatively spliced [human, Namalwa cells, Peptide Partial, 391 aa] transcription factor LSF [Homo sapiens]	51618	58	9
MCM2	**FIP2**	**gi|434753 gi|33356547**	minichromosome maintenance complex component 2 [Homo sapiens]	102516	57	4
EXOC4	**FIP2**	**gi|12697943 gi|82546830**	Exoc5 in Y2H list. exocyst complex component 4 [Homo sapiens]	111170	57	3
MPG	**FIP5**	**gi|178654**	alkyl-N-purine-DNA glycosylasee [Homo sapiens]	25777	57	3
hMre11	**FIP2**	**gi|3912938**	MRE11 homologue hMre11 [Homo sapiens]	77908	55	7
OSBPL9	**FIP2**	**gi|14042588 gi|22547165**	oxysterol binding protein-like 9 isoform a [Homo sapiens]	63059	55	2
LMAN2	**FIP5**	**gi|5803023**	lectin, mannose-binding 2 precursor [Homo sapiens]	40545	55	1
KIF4A	**FIP1**	**gi|5802957**	kinesin family member 4A [Homo sapiens]	141402	54	22
PRMT5	**2C**	**gi|20070220**	in Y2H is PRMT10. protein arginine methyltransferase 5 isoform a [Homo sapiens]	73322	54	6
GTF2IRD1	**FIP2**	**gi|10047171 gi|122937295**	BEN domain containing 3 [Homo sapiens]	95556	54	3
SHCBP1	**FIP1**	**gi|7661950**	SH2 domain binding protein 1 [Homo sapiens]	134332	54	1
USP10	**FIP3**	**gi|3098601**	Ras-GAP SH3 binding protein [Homo sapiens]	50795	54	1
LRRC59	**FIP5**	**gi|40254924**	leucine rich repeat containing 59 [Homo sapiens]	35308	53	7
RUVBL2	**FIP4**	**gi|5730023**	RuvB-like 2 [Homo sapiens]	51296	53	3
ORC5L	**FIP4**	**gi|3033526**	origin recognition complex subunit 5 [Homo sapiens]	50864	53	2
OTUD4	**SC1**	**gi|118572680**	RecName: Full=OTU domain-containing protein 4; AltName: Full=HIV-1-induced protein HIN-1	124823	52	6
UHRF1	**FIP2**	**gi|6815251 gi|116138198**	Ubiquitin-like with PHD and ring finger domains 1 [Homo sapiens]	92598	51	6
GTF3C4	**FIP2**	**gi|6175593 gi|156119605**	general transcription factor IIIC 4 [Homo sapiens]	93235	51	4
TP53	**FIP4**	**gi|189479**	p53 cellular tumor antigen [Homo sapiens] double hits **gi|116668388** Chain A, Human P53 Core Domain Mutant M133l-V203a-N239y-N268d-R282w	44177	51	3
DPM1	**FIP5**	**gi|2258418**	dolichol monophosphate mannose synthase [Homo sapiens]	28857	51	1
EXOSC7	**FIP5**	**gi|473949**	Exoc5 in Y2H list. exosome component 7 [Homo sapiens]	32428	51	1
DNM3	**FIP2**	**gi|42544243**	dynamin 3 isoform a [Homo sapiens]	97434	50	10

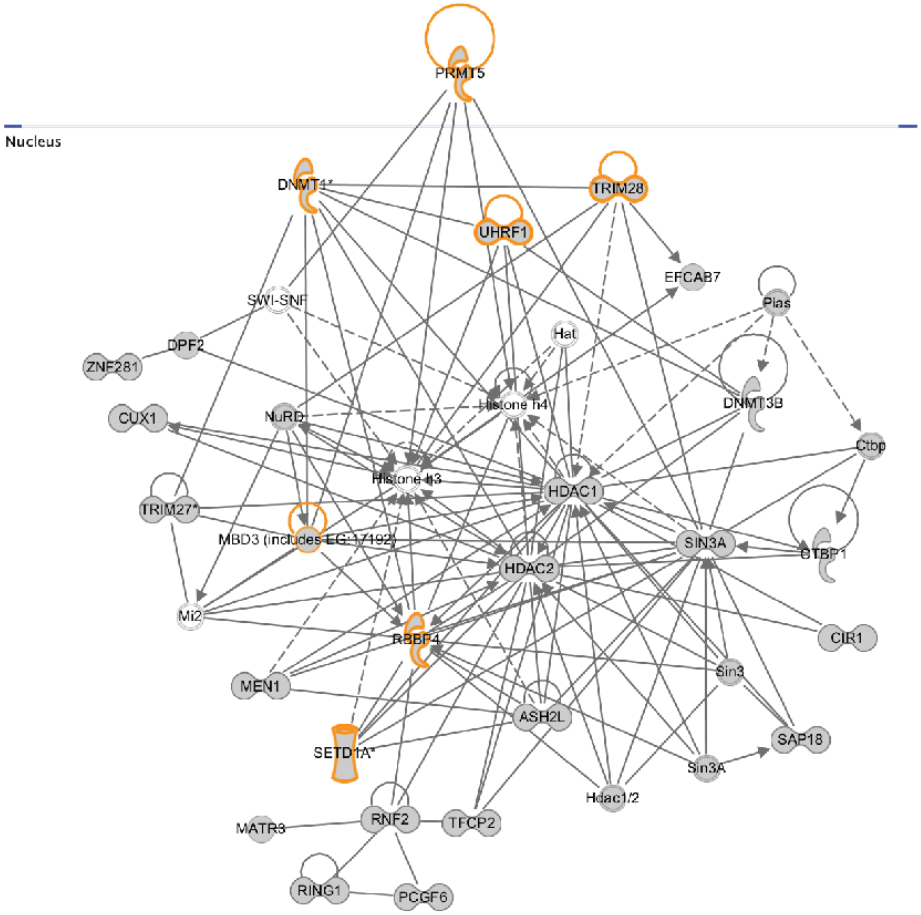
DNM3	**FIP2**	**gi|20521666 gi|209915561**	dynamin 3 isoform b [Homo sapiens]	97021	50	9
ATP1B1	**FIP4**	**gi|4502281**	in Y2H is ATP1b2. Na+/K+ -ATPase beta 3 subunit [Homo sapiens]	31834	50	1
COPG	**FIP2**	**gi|7023756 gi|11559929**	coatomer protein complex, subunit gamma 1 [Homo sapiens]	98967	49	1
PPPIA	**FIP5**	**gi|190281**	protein phosphatase I alpha subunit (PPPIA) (EC 3.1.3.16) [Homo sapiens]	35844	48	2
CSNK2A1	**FIP4**	**gi|599778**	casein kinase II alpha subunit [Homo sapiens]	45242	48	1
STOML2	**FIP4**	**gi|2984585**	stomatin (EPB72)-like 2 [Homo sapiens]	38624	47	3
DEK	**FIP4**	**gi|4503249**	DEK oncogene isoform 1 [Homo sapiens]	42933	46	6
PSPC1	**FIP3**	**gi|7023323**	paraspeckle protein 1 [Homo sapiens]	58820	45	5
GNAS	**FIP4**	**gi|4504047**	GNAS complex locus GNASL [Homo sapiens]		45	2
PICALM	**FIP3**	**gi|1373146**	phosphatidylinositol binding clathrin assembly protein [Homo sapiens]	70879	45	2
CDK2	**FIP5**	**gi|34809859**	Chain A, Cdk2 In Complex With A Disubstituted 2, 4-Bis Anilino Pyrimidine Cdk4 Inhibitor	34062	44	6
DNAH5	**FIP4**	**gi|19115954**	dynein, axonemal, heavy chain 5 [Homo sapiens]	532504	44	5
FAT2	**FIP4**	**gi|13787217**	FAT tumor suppressor 2 precursor [Homo sapiens]		44	5
XPNPEP3	**FIP2**	**gi|11559925**	X-prolyl aminopeptidase (aminopeptidase P) 3, putative [Homo sapiens]	57624	44	4
CHGB	**FIP4**	**gi|134461**	RecName: Full=Secretogranin-1; AltName: Full=Secretogranin I; Short=SgI; AltName: Full=Chromogranin-B; Short=CgB; Contains: RecName: Full=GAWK peptide; Contains: RecName: Full=CCB peptide; Flags: Precursor		44	3
DNAH8	**FIP4**	**gi|14335444**	axonemal dynein heavy chain 8 [Homo sapiens]	518134	44	3
ZC3H8	**FIP4**	**gi|21618936**	ZC3H8 protein [Homo sapiens]		44	3
CLPX	**FIP3**	**gi|7242140**	ClpX caseinolytic protease X homolog precursor [Homo sapiens]	69922	44	2
LAS1L	**FIP2**	**gi|13654270**	LAS1-like isoform 1 [Homo sapiens]	83982	44	2
STRAP	**FIP4**	**gi|4519417**	WD-40 repeat protein [Homo sapiens] unr-interacting protein [Homo sapiens]serine-threonine kinase receptor-associated protein [Homo sapiens]	38814	44	2
HADHB	**FIP4**	**gi|4504327**	mitochondrial trifunctional protein, beta subunit precursor [Homo sapiens]	51547	43	4
MAPRE2	**FIP4**	**gi|10346135**	microtubule-associated protein, RP/EB family, member 2 isoform 1 [Homo sapiens]	37236	43	4
MARS	**FIP2**	**gi|1702932**	yeast methionyl-tRNA synthetase homolog [Homo sapiens]	102262	42	5
PTBP1	**FIP3**	**gi|4506243**	polypyrimidine tract-binding protein 1 isoform a [Homo sapiens]	59767	42	3
ORC4L	**FIP4**	**gi|2736149**	putative replication initiator origin recognition complex subunit Orc4Lp [Homo sapiens]origin recognition complex subunit 4	50717	42	2
ANXA6	**FIP3**	**gi|35218**	calelectrin [Homo sapiens]	76199	42	1
LBP1a	**FIP3**	**gi|545524**	LBP-1a=transcription factor binding to initiation site of HIV-1 alternatively spliced [human, Namalwa cells, Peptide, 504 aa]	56846	41	5
SRPRB	**FIP5**	**gi|284795266**	signal recognition particle receptor, beta subunit [Homo sapiens]	29912	41	3
DAP3	**FIP4**	**gi|4758118**	death-associated protein 3 [Homo sapiens]	45880	41	2
H2A.2	**2F**	**gi|31979**	In control is histone cluster 1(H1d), histone H2A.2 [Homo sapiens]	13898	41	1
LARP4	**FIP2**	**gi|15053987**	c-Mpl binding protein [Homo sapiens]La ribonucleoprotein domain family, member 4	42820	41	1
RBBP4	**FIP4**	**gi|5032027**	retinoblastoma binding protein 4 isoform a [Homo sapiens]	47911	40	13
CTNNBL1	**FIP3**	**gi|18644734**	beta catenin-like 1 [Homo sapiens]	65588	40	4
DPF2	**FIP4**	**gi|5454004**	D4, zinc and double PHD fingers family 2 [Homo sapiens]	45268	40	2
GCN1L1	**FIP1**	**gi|3970973 gi|54607053**	GCN1 general control of amino-acid synthesis 1-like 1 [Homo sapiens]	294919	40	2
LEPRE1	**FIP2**	**gi|11127636 gi|119627549**	leucine proline-enriched proteoglycan (leprecan) 1, isoform CRA_c	91681	40	2
FUBP3	**FIP3**	**gi|1575609**	In control is fuse binding protein 2. FUSE binding protein 3 [Homo sapiens]far upstream element (FUSE) binding protein 3	64239	40	1
PWP2	**FIP2**	**gi|1438062**	PWP2 periodic tryptophan protein homolog (yeast) [Homo sapiens]	103342	39	6
Setd1a	**FIP1**	**gi|6683126 gi|55741677**	SET domain containing 1A [Homo sapiens]	186918	39	3
CSDE1	**FIP2**	**gi|3387902 gi|16356661**	NRAS-related protein [Homo sapiens]	80046	39	2
NOP56	**FIP3**	**gi|2230878**	NOP56 ribonucleoprotein homolog (yeast) NOP56 ribonucleoprotein homolog (yeast) [Homo sapiens]	67206	39	1
SMC3	**FIP1**	**gi|4885399**	chondroitin sulfate proteoglycan 6 (bamacan), isoform CRA_b [Homo sapiens]	141853	38	8

LIG3	**FIP2**	**gi|860963**	DNA ligase III [Homo sapiens]	103937	38	6
PPP2R4	**FIP3**	**gi|178663**	medium tumor antigen-associated 61-kD protein [Homo sapiens] phosphatase 2A regulatory subunit [Homo sapiens]	65980	38	4
EXOSC10	**FIP2**	**gi|4505917**	Exoc5 in Y2H list. exosome component 10 isoform 2 [Homo sapiens]	98826	38	3
MED4	**FIP1**	**gi|4826626 gi|4884564**	vitamin D3 receptor interacting protein [Homo sapiens]mediator complex subunit 4	158212	38	3
MARCKSL1	**FIP5**	**gi|13491174**	MARCKS in Y2H. MARCKS-like 1 [Homo sapiens]	19574	38	1
IDH3A	**FIP5**	**gi|5031777**	isocitrate dehydrogenase 3 (NAD+) alpha precursor [Homo sapiens]	12397	37	16
IFT172	**FIP3**	**gi|46358428**	selective LIM binding factor homolog [Homo sapiens] intraflagellar transport 172 homolog (Chlamydomonas), isoform CRA_a [Homo sapiens]	199175	37	9
PSMD3	**FIP3**	**gi|2656092**	proteasome subunit p58 [Homo sapiens]proteasome (prosome, macropain) 26S subunit, non-ATPase, 3 [Homo sapiens]	61082	37	3
CCNB1	**FIP4**	**gi|14327896**	cyclin B1 [Homo sapiens]	48591	37	1
RBM10	**FIP2**	**gi|34785044**	RBM10 protein [Homo sapiens]	59274	37	1
ABCF2	**FIP3**	**gi|3005931**	ABC transporter [Homo sapiens] ATP-binding cassette, sub-family F, member 2 isoform a [Homo sapiens]	71815	36	4

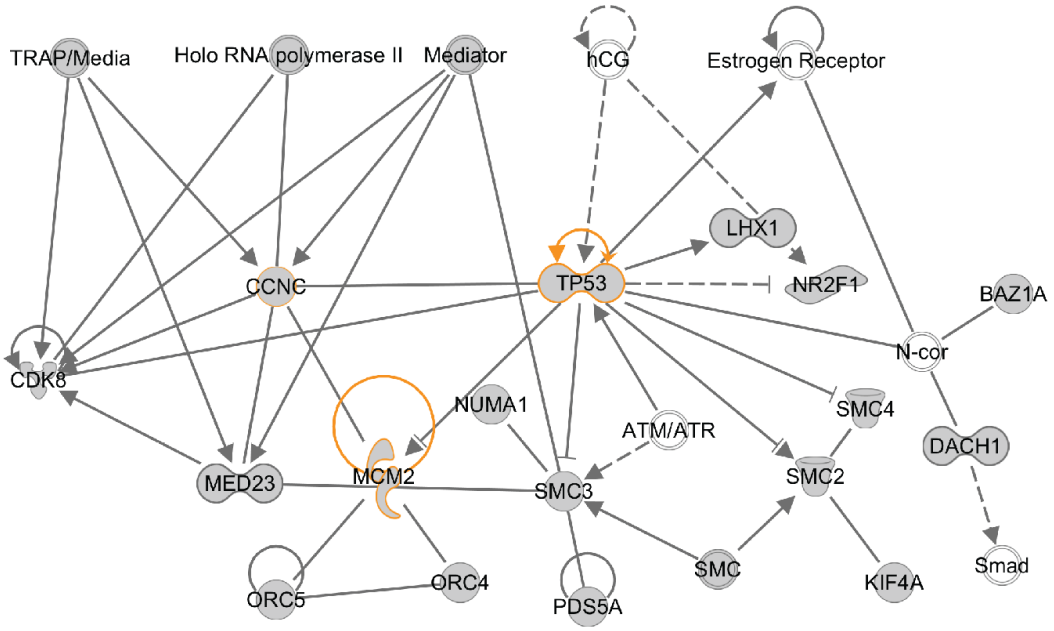
A.



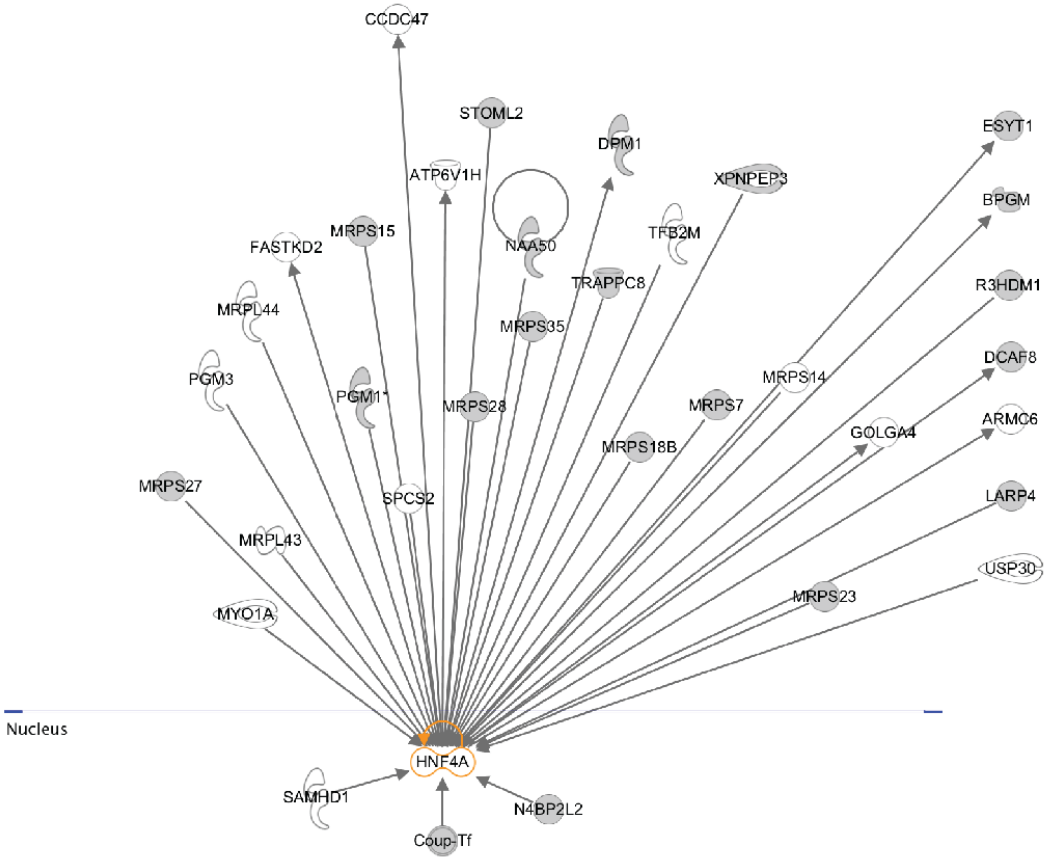
B.


PRMT5
Nucleus
DNMT4*
TRIM28
UHRF1
EFCAB7
Plas
SWI-SNF
Hat
DPF2
ZNF281
Histone h4
DNMT3B
CUX1
NuRD
Ctbp
Histone h3
HDAC1
TRIM27*
MBD3 (includes EG:17192)
SIN3A
BTBP1
HDAC2
Mi2
RBBP4
CIR1
Sin3
MEN1
ASH2L
SETD1A*
SAP18
Sin3A
Hdac1/2
MATR3
RNF2
TFCP2
RING1
PCGF6

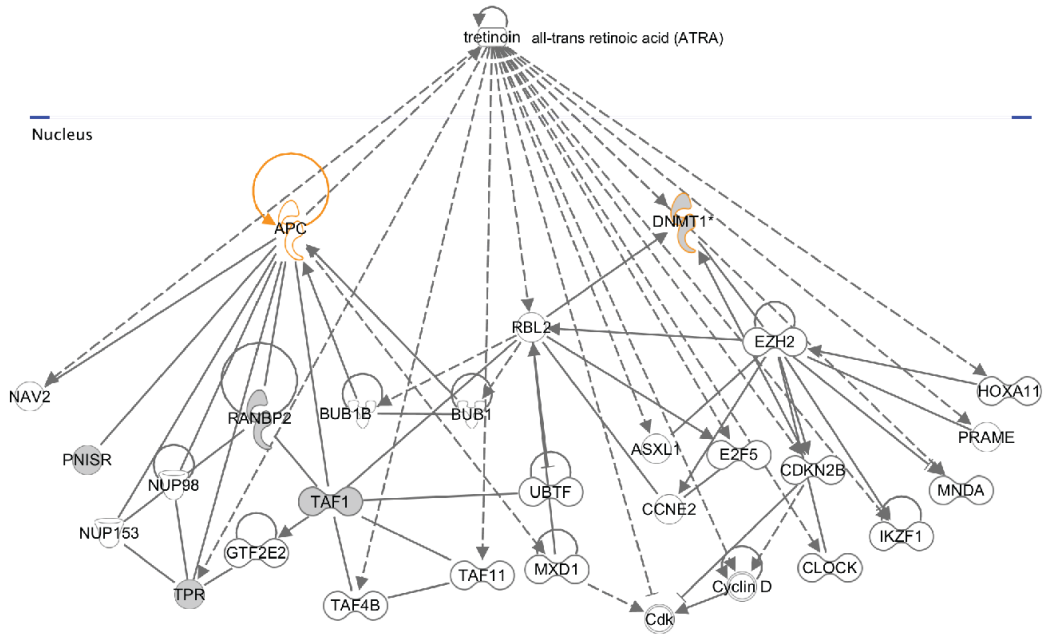
C.

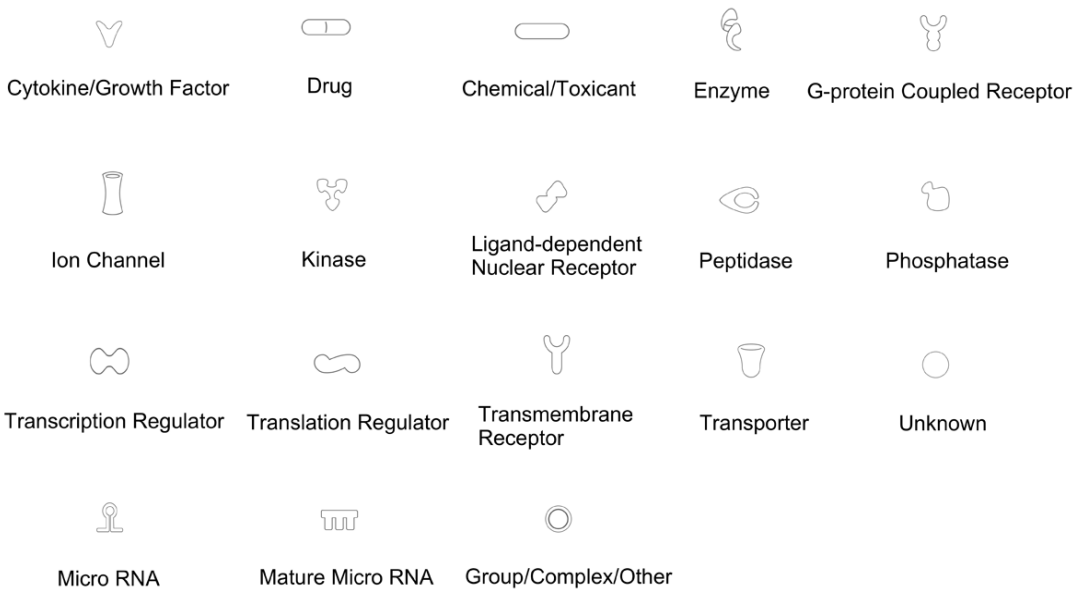

TRAP/Media
Holo RNA polymerase II
Mediator
hCG
Estrogen Receptor
LHX1
CCNC
TP53
NR2F1
BAZ1A
N-cor
CDK8
NUMA1
SMC4
ATM/ATR
DACH1
MED23
MCM2
SMC3
SMC2
SMC
Smad
ORC5
ORC4
PDS5A
KIF4A

D.


CCDC47
STOML2
DPM1
ESYT1
ATP6V1H
XPNPEP3
BPGM
MRPS15
FASTKD2
NAA50
TFB2M
TRAPPC8
R3HDM1
MRPL44
MRPS35
DCAF8
PGM3
PGM1
MRPS28
MRPS14
MRPS7
ARMC6
MRPS18B
GOLGA4
MRPS27
SPCS2
LARP4
MRPL43
USP30
MRPS23
MYO1A
Nucleus
HNF4A
SAMHD1
N4BP2L2
Coup-Tf

E.


tretinoin
all-trans retinoic acid (ATRA)
Nucleus
APC
DNMT1*
RBL2
EZH2
NAV2
HOXA11
RANBP2
BUB1B
BUB1
PRAME
PNISR
ASXL1
E2F5
CDKN2B
NUP98
UBTF
MNDA
TAF1
CCNE2
NUP153
IKZF1
GTF2E2
MXD1
CLOCK
TAF11
Cyclin D
TPR
TAF4B
Cdk

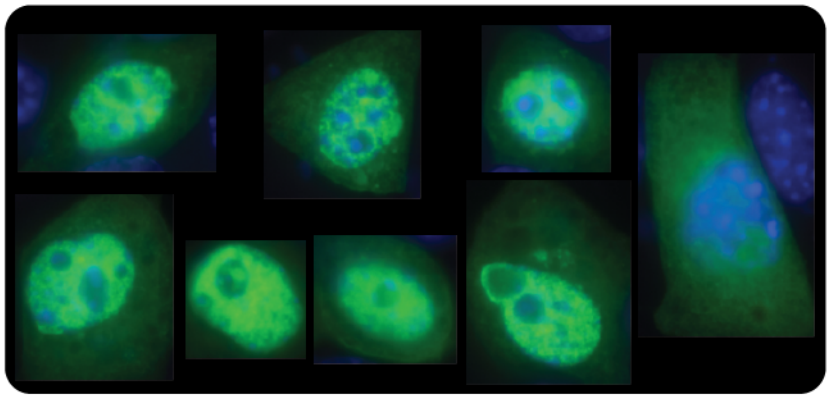


Supplementary Figure. 3.1. Significant protein clusters from the IPA network analysis. (A) DNA repair, recombination, and replication. (B) Histone modification, and chromatin remodeling. (C) Lipid Metabolism, and Endocrine System Development and Function. (D) Genetic Disorder, Respiratory Disease, and Lipid Metabolism (Orphan nuclear receptor) (E) Cellular Growth and Proliferation (RXR activation, retinoic acid associated, much less significant). Note the proteins with gray filling are the candidates from my screens; the ones without gray filling are those involved in the indicated pathway but not in my candidate list. The solid lines are direct relationships; and the dotted lines are indirect relationships. The arrows generally indicate the positive regulation.

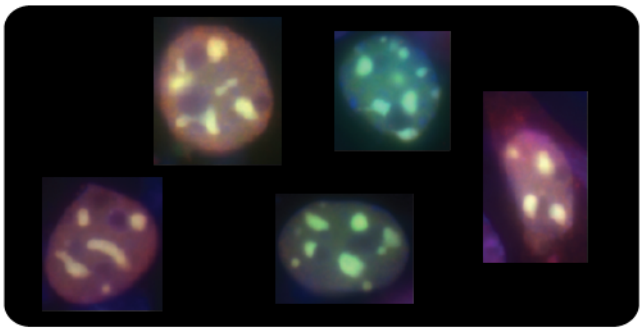
Supplementary Figures for Chapter 4.

More pictures for each IF microscopy of single and double transfections of Mbd4 and its protein association candidates are shown. The cell sizes between single and double transfections of Mbd4 with respective protein candidate were zoomed in at the same proportion so that the changes of nucleus size can be observed.

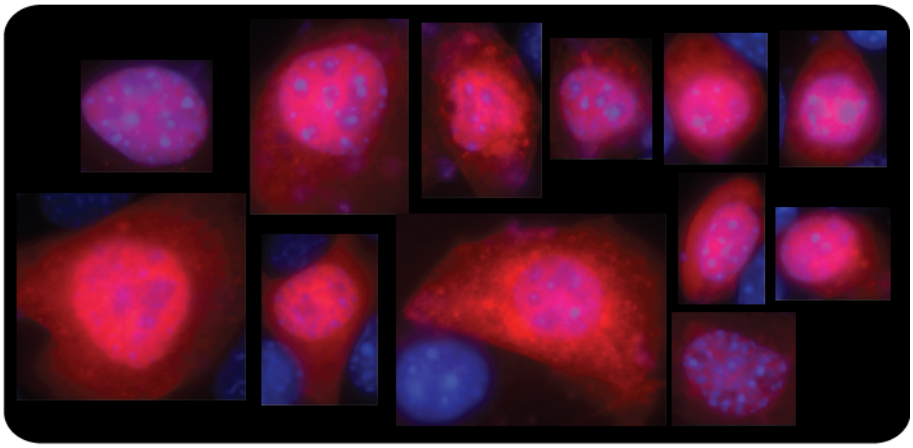
A1. GFP-Mlh1 single transfection + DAPI



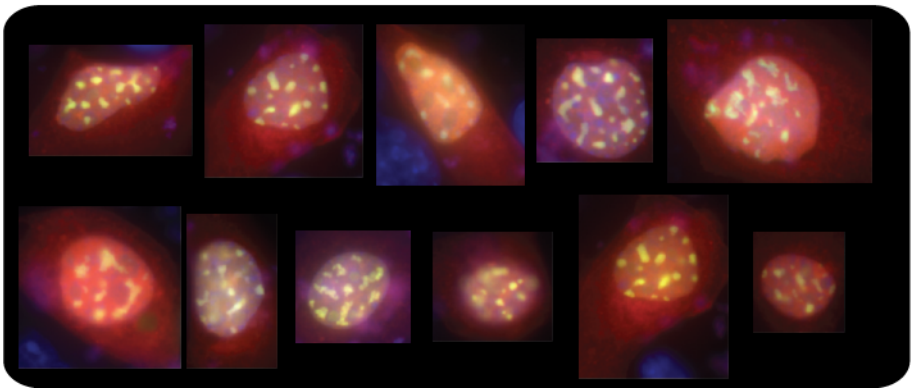
A2. GFP-Mlh1 + FLAG-hMBD4 +DAPI



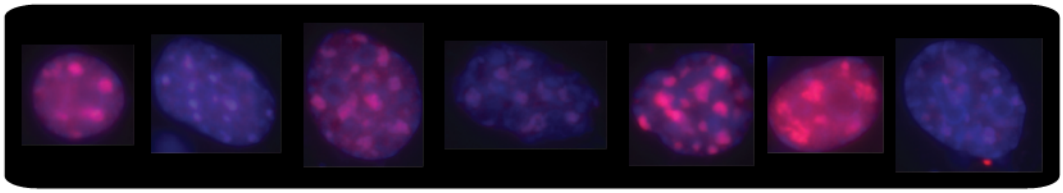
B1. Mcherry-USP7 single transfection + DAPI



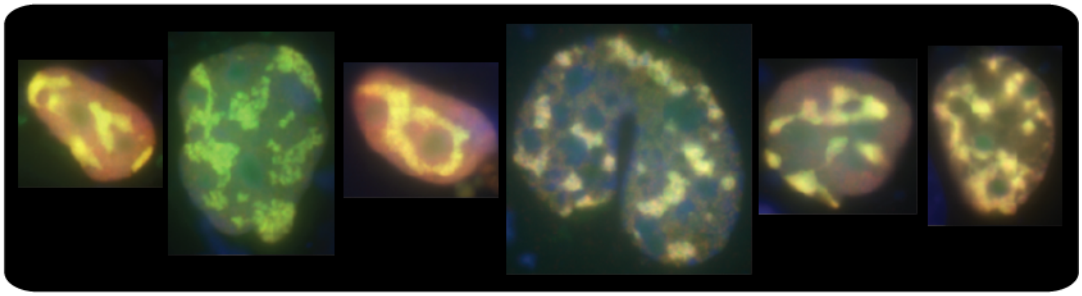
B2. Mcherry-USP7 + GFP-mMbd4 + DAPI



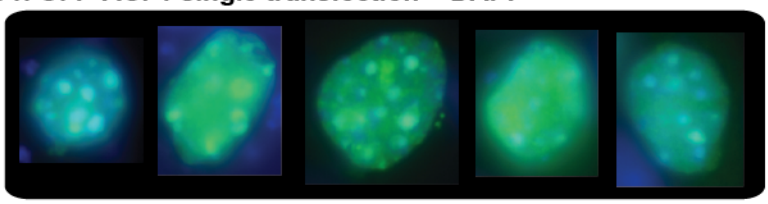
C1. FLAG-UHRF1 single transfection + DAPI



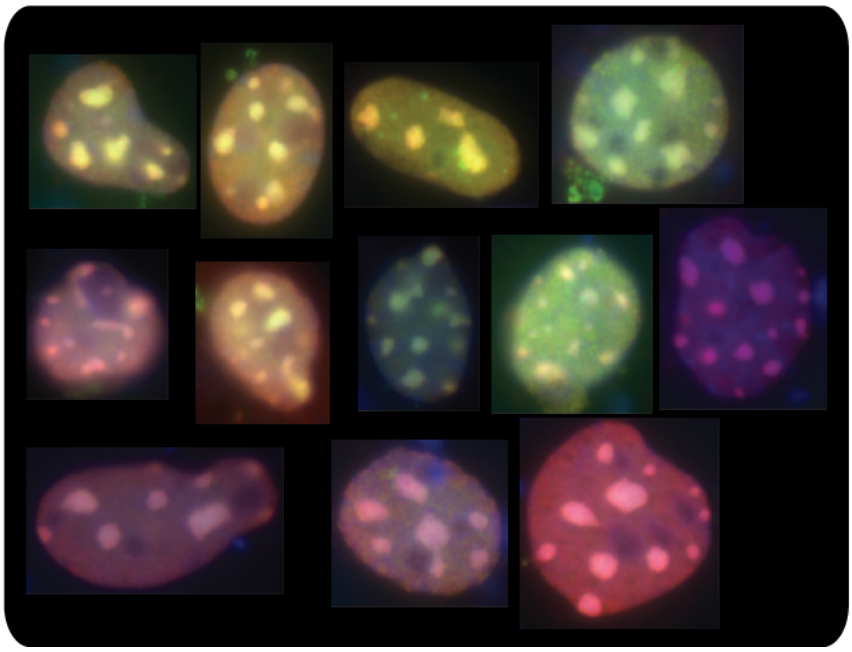
C2. FLAG-UHRF1 + GFP-mMbd4 +DAPI



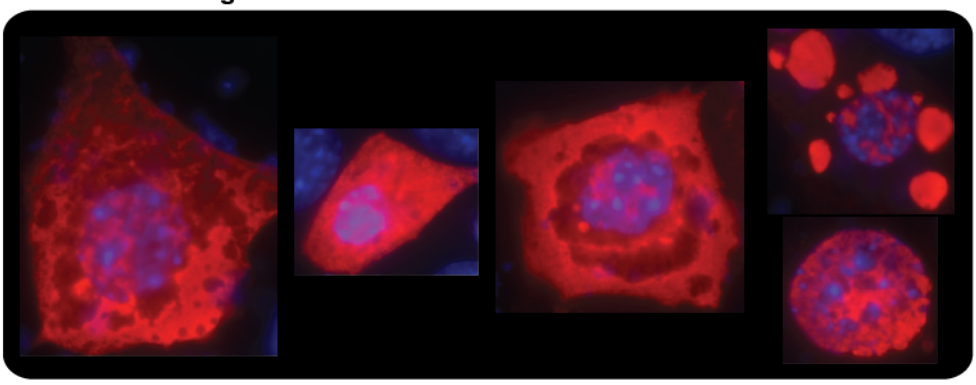
D1. GFP-ACF1 single transfection + DAPI



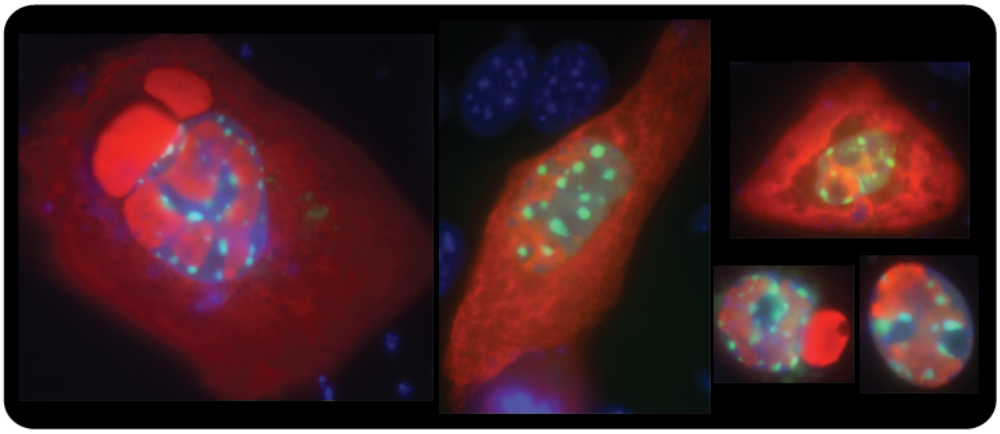
D2. GFP-ACF1 + FLAG-hMBD4 + DAPI



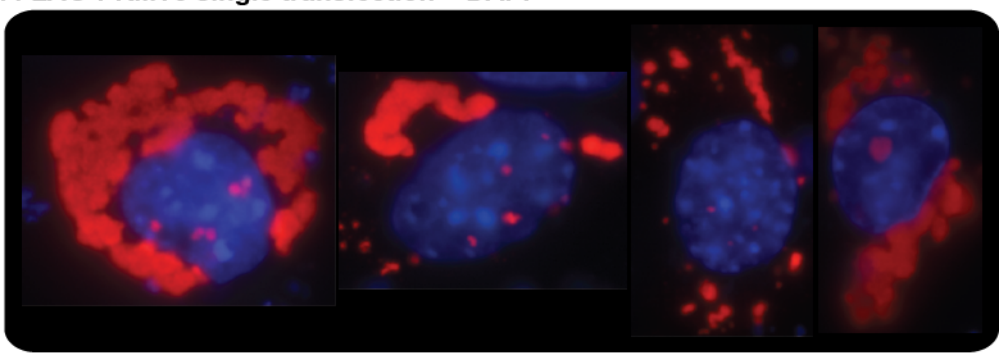
E1. FLAG-KAP1 single transfection + DAPI



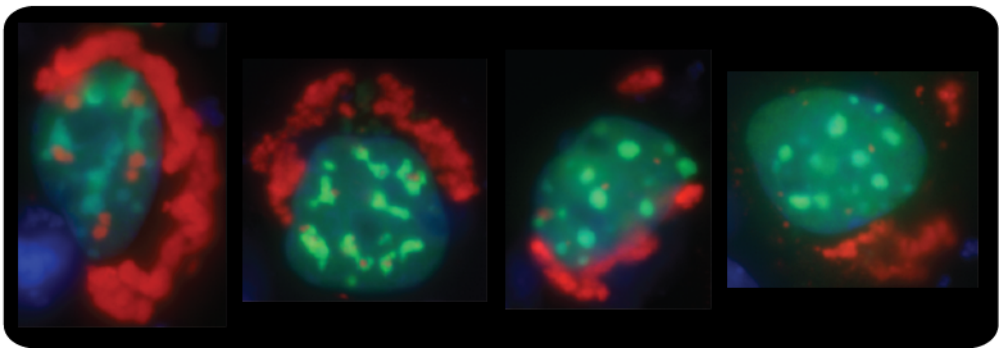
E2. FLAG-KAP1 + GFP-mMBD4 + DAPI



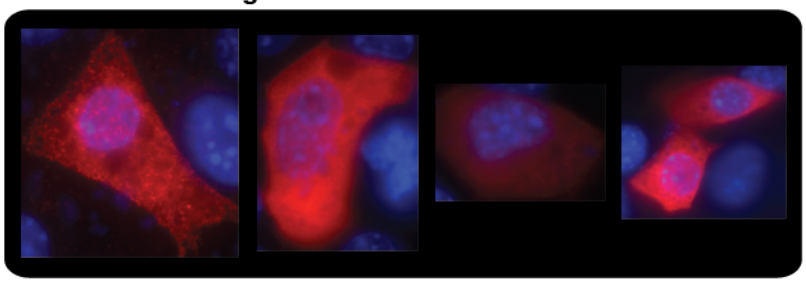
F1. FLAG-PRMT5 single transfection + DAPI



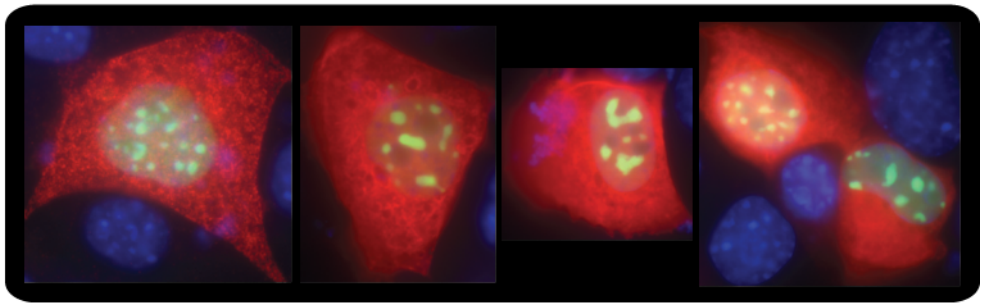
F2. FLAG-PRMT5 + GFP-mMbd4 + DAPI



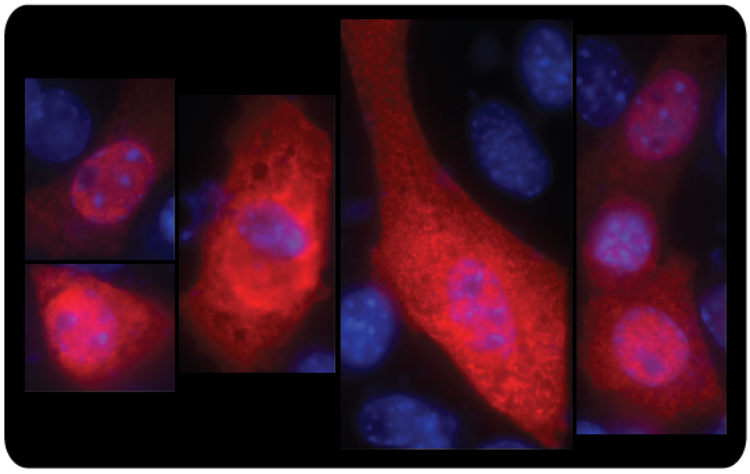
G1. FLAG-PRMT3 single transfection + DAPI



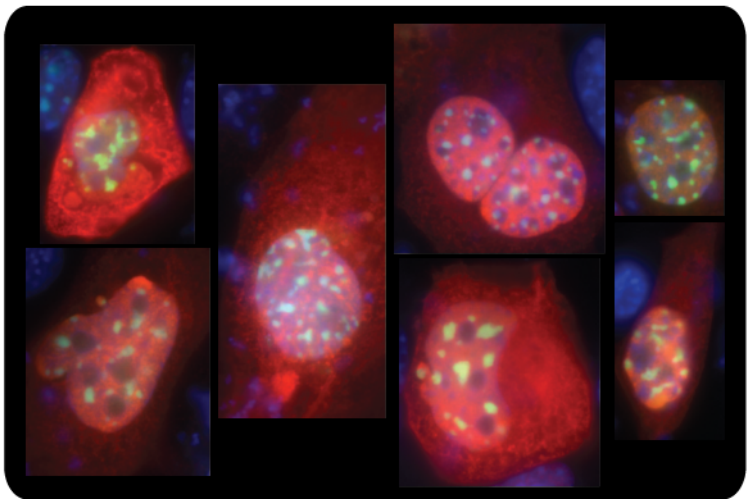
G2. FLAG-PRMT3 + GFP-mMbd4 + DAPI



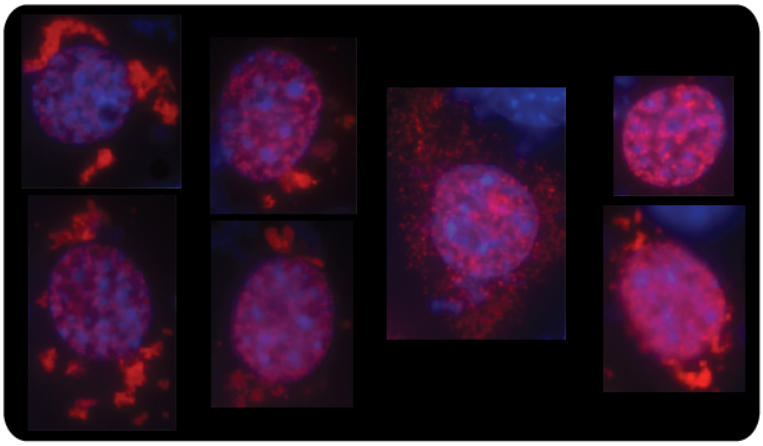
H1. FLAG-PRMT1 single transfection + DAPI



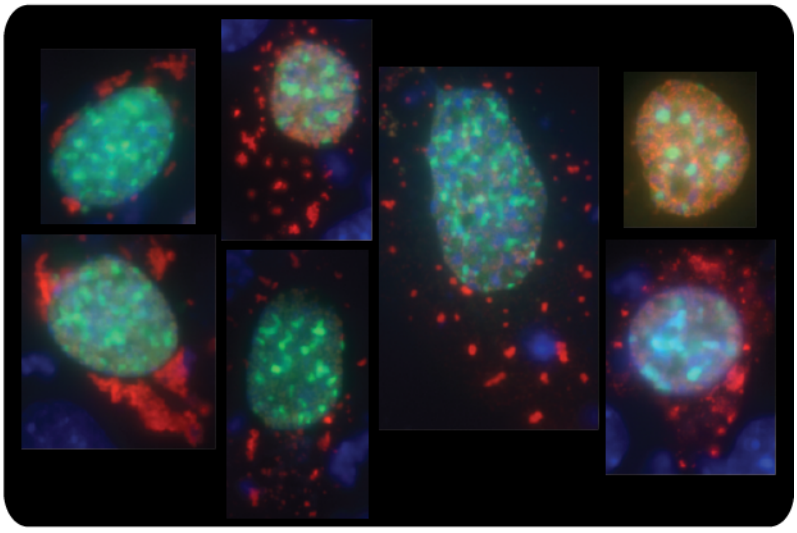
H2. FLAG-PRMT1 + GFP-mMbd4 + DAPI



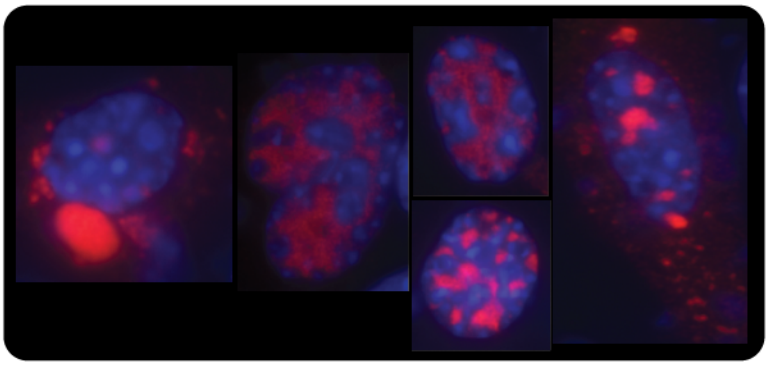
I1. FLAG-CFP1 single transfection + DAPI



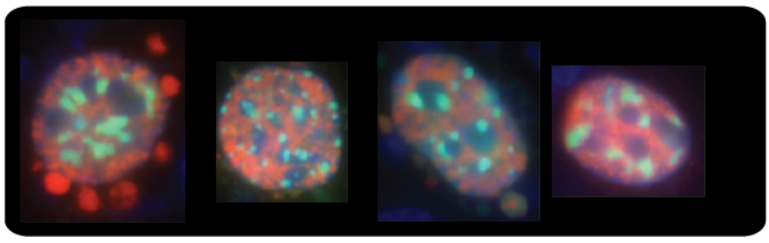
I2. FLAG-CFP1 + GFP-mMBD4 + DAPI



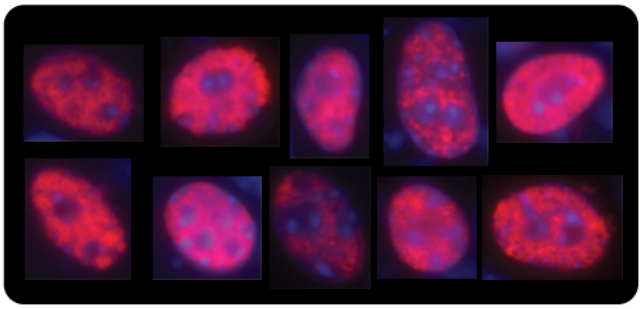
J1. FLAG-Setd1a single transfection + DAPI



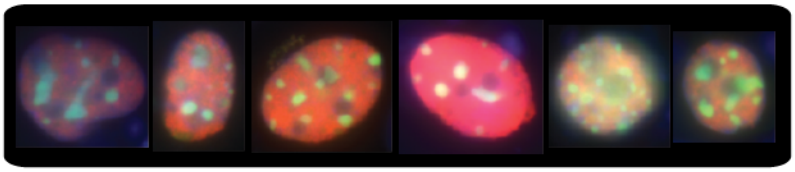
J2. FLAG-Setd1a + GFP-mMbd4 + DAPI



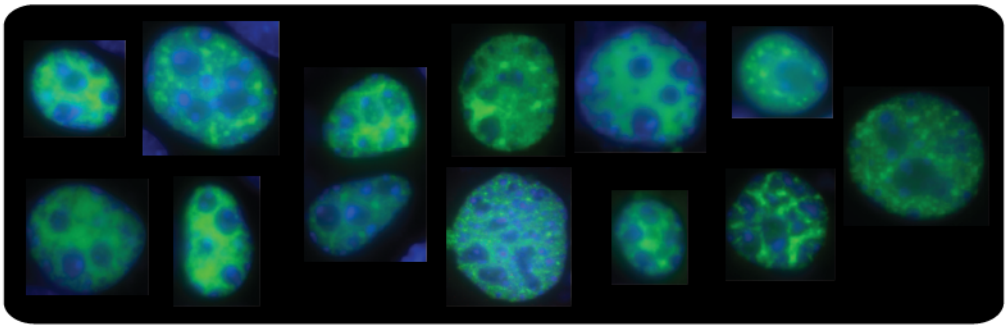
K1. FLAG-DEK single transfection + DAPI



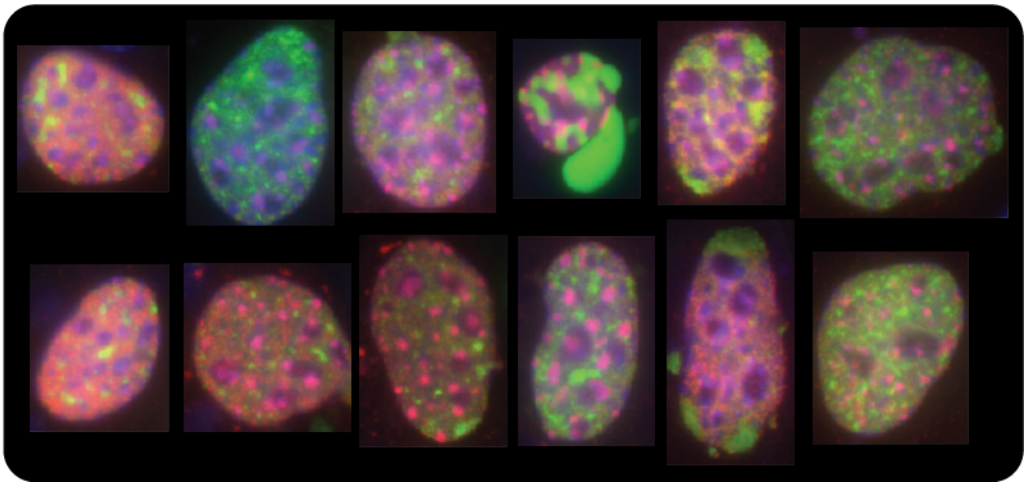
K2. FLAG-DEK + GFP-mMbd4 + DAPI



L1. GFP-SC35 single transfection + DAPI



L2. GFP-SC35 + FLAG-hMBD4 + DAPI



Supplementary Table 5.1 Transcripts significantly misregulated in virus transformed Mbd4$^{-/-}$ cells (vtKO). Shown are all transcripts mis-regulated >2-fold in Mbd4$^{-/-}$ cells (relative to wildtype cells) and their average signals on the BeadChips. Positive number in the fold change column indicates upregulation in knockout cells; negative number means downregulation.

Gene Name	Wildtype_AVG_Signal	Knockout_AVG_Signal	Fold Change	SEARCH_KEY
LOC666403	232.6723	6684.256	28.7	ILMN_199173
2310014G06Rik	39.19115	895.7907	22.9	ILMN_187083
Slc1a3	148.3591	3347.447	22.6	ILMN_213165
Slc2a10	8.122583	119.9356	14.8	ILMN_211721
Ebf3	36.31731	460.7043	12.7	ILMN_184911
Sncg	13.27798	154.2386	11.6	ILMN_209715
Igf2	288.1348	3343.976	11.6	ILMN_209642
Enpp1	11.85449	136.6972	11.5	ILMN_210635
Gpx7	19.21198	218.149	11.4	ILMN_220603
Prss35	24.10223	268.3559	11.1	ILMN_210857
Igfbp5	28.2518	314.4135	11.1	ILMN_190841
scl0002702.1_3805	44.66982	482.2918	10.8	ILMN_195541
Serpine2	248.5313	2652.274	10.7	ILMN_204512
AI467606	11.2887	107.6597	9.5	ILMN_220903
Chst7	11.97757	108.344	9.0	ILMN_211189
Mmp23	26.13883	236.3173	9.0	ILMN_215522
Ccdc3	15.60841	139.2129	8.9	ILMN_210971
Kcnab2	9.48685	84.29489	8.9	ILMN_220644
Cln3	17.22556	148.4376	8.6	ILMN_220100
Pygo1	13.37353	112.8549	8.4	ILMN_217089
H19	165.1176	1341.726	8.1	ILMN_219881
Flywch2	14.01056	110.8243	7.9	ILMN_213672
Col2a1	33.20084	248.082	7.5	ILMN_188506
Lrrc17	23.63244	173.2555	7.3	ILMN_194980
LOC381230	918.942	6690.498	7.3	ILMN_198449
Clip4	16.34958	112.6116	6.9	ILMN_193898
Tspyl3	13.15097	90.25643	6.9	ILMN_196361
Slc2a3	8.877837	59.52139	6.7	ILMN_211512
Rgs16	44.74363	291.8918	6.5	ILMN_209950
BC011487	20.86034	135.8858	6.5	ILMN_211661
Mest	33.584	216.8244	6.5	ILMN_212378
Foxc1	87.86136	565.3849	6.4	ILMN_214651
Diras2	19.5947	124.2857	6.3	ILMN_194373
Irx3	105.5247	664.3024	6.3	ILMN_224225
Stmn2	17.19265	105.9531	6.2	ILMN_210469
Cryl1	44.49581	271.6955	6.1	ILMN_217939
Sdpr	300.0698	1805.715	6.0	ILMN_217783
C1qtnf3	228.1418	1352.817	5.9	ILMN_209151
Larp6	13.91039	79.45901	5.7	ILMN_209330
LOC100045780	61.27566	341.5616	5.6	ILMN_223951
Runx1t1	40.25037	215.7801	5.4	ILMN_210856
Adamtsl1	12.63175	67.65624	5.4	ILMN_219541
Rab15	60.80197	324.8687	5.3	ILMN_218127
Nfatc4	181.6671	960.3871	5.3	ILMN_214372
E330009F12Rik	16.25993	83.07156	5.1	ILMN_206616
Usp13	10.46358	53.01886	5.1	ILMN_184877
Gata2	147.2167	726.5109	4.9	ILMN_211079
Unc5c	29.39423	144.2213	4.9	ILMN_188618
Pcdh17	16.36931	80.18647	4.9	ILMN_236878
Bglap2	20.78012	100.8721	4.9	ILMN_226483

Tnfsf11	16.5336	79.48047	4.8	ILMN_194367
Capn6	279.4984	1335.438	4.8	ILMN_218388
Ldoc1l	23.38252	110.4228	4.7	ILMN_209512
Osbpl6	9.936504	46.62185	4.7	ILMN_190013
Creb5	12.42477	58.25422	4.7	ILMN_217830
LOC100048376	16.8174	78.63554	4.7	ILMN_206536
Hgf	6.179766	28.1847	4.6	ILMN_220957
Arc	16.43672	74.95475	4.6	ILMN_209647
Magohb	22.80122	102.447	4.5	ILMN_216019
Aoc3	15.15454	67.85607	4.5	ILMN_203969
Dleu2	21.33905	95.08724	4.5	ILMN_184991
1810031K17Rik	12.81047	57.03035	4.5	ILMN_209092
Pappa	17.03749	75.28543	4.4	ILMN_212814
Sgsm1	16.92474	74.51131	4.4	ILMN_218133
D130017D19Rik	40.73962	179.0964	4.4	ILMN_206166
Thg1l	35.37402	154.4858	4.4	ILMN_246881
Sorcs2	24.21794	104.6961	4.3	ILMN_218384
D130032J17Rik	39.63448	170.6335	4.3	ILMN_205501
Arsj	21.59859	91.52135	4.2	ILMN_217748
Aatk	13.00303	54.81442	4.2	ILMN_221321
Hebp1	50.2327	211.4887	4.2	ILMN_218714
Fxyd5	93.2799	392.0496	4.2	ILMN_221977
Tuba4a	9.227804	38.28577	4.1	ILMN_195615
2410075B13Rik	12.92626	53.01389	4.1	ILMN_217814
Fez1	136.397	553.4005	4.1	ILMN_214033
Limch1	19.47171	78.83987	4.0	ILMN_212627
Acta1	8.152735	32.77879	4.0	ILMN_221671
B4galnt1	34.55429	138.5764	4.0	ILMN_220892
Stag3	21.0913	84.57735	4.0	ILMN_212542
Itga11	62.65017	250.8298	4.0	ILMN_210182
C030003D03Rik	10.38529	41.37816	4.0	ILMN_184855
Flrt2	30.90172	122.905	4.0	ILMN_215172
Apol7a	11.07717	43.34136	3.9	ILMN_211553
2310046K01Rik	11.45495	44.0897	3.8	ILMN_216823
6330406I15Rik	313.1269	1202.792	3.8	ILMN_213942
Slc17a5	49.25476	188.8229	3.8	ILMN_222093
AI747699	32.90839	125.62	3.8	ILMN_258552
Foxg1	49.37395	187.9115	3.8	ILMN_209777
Anxa4	86.26463	327.9602	3.8	ILMN_223968
Aqp1	110.8348	420.5979	3.8	ILMN_216314
Cap1	313.7124	1185.877	3.8	ILMN_214505
Pdgfra	236.6705	887.0174	3.7	ILMN_212227
Mlf1	57.1548	213.8385	3.7	ILMN_246637
D730042P09Rik	34.11002	127.5683	3.7	ILMN_202674
Cbr2	78.7692	294.0188	3.7	ILMN_222912
Farsa	19.9718	74.40858	3.7	ILMN_184453
Lrp11	44.14596	164.1566	3.7	ILMN_216574
Tro	17.24827	63.76163	3.7	ILMN_227694
Scara3	513.7833	1897.084	3.7	ILMN_219254
Cpe	1367.751	5022.34	3.7	ILMN_223071
5430433G21Rik	22.12658	81.14508	3.7	ILMN_220885
Mcf2l	8.850212	32.45078	3.7	ILMN_207033
1810062O18Rik	15.32137	55.84475	3.6	ILMN_192867
Endod1	55.65228	202.0621	3.6	ILMN_221277
Ripk4	17.5804	63.78962	3.6	ILMN_231082
Gstm4	9.989587	36.12521	3.6	ILMN_220200
LOC665032	1900.742	6861.591	3.6	ILMN_196998
1810014F10Rik	24.02255	85.23515	3.5	ILMN_209922
Fabp3	75.98491	269.3723	3.5	ILMN_210798
Rnf130	289.5708	1022.093	3.5	ILMN_218606
Ddr2	16.0882	56.76863	3.5	ILMN_217700

Thsd1	9.372628	33.04095	3.5	ILMN_192069
1600021P15Rik	90.49725	318.0388	3.5	ILMN_185891
Fahd2a	13.73765	48.05991	3.5	ILMN_210830
A730096C04Rik	34.90505	121.8205	3.5	ILMN_204109
3830406C13Rik	46.446	161.0942	3.5	ILMN_207451
A730020E08Rik	7.588012	26.14706	3.4	ILMN_192463
6230400G14Rik	68.36479	235.4947	3.4	ILMN_193477
Ocel1	15.97345	54.55274	3.4	ILMN_216109
Ppbp	32.97945	111.7722	3.4	ILMN_209889
Prkd1	15.0822	50.47045	3.3	ILMN_214388
AI428936	53.84797	179.8416	3.3	ILMN_211797
Mrgprf	59.72216	198.023	3.3	ILMN_223171
Cpxm1	56.12373	185.0691	3.3	ILMN_214237
4833431P16Rik	41.60565	136.9423	3.3	ILMN_204798
scl0002540.1_6	19.93278	65.41423	3.3	ILMN_196212
Ly6a	630.8228	2052.006	3.3	ILMN_209150
1500009L16Rik	38.73602	125.9192	3.3	ILMN_187731
AA467197	10.92054	35.3493	3.2	ILMN_236731
Hbegf	198.2778	641.2649	3.2	ILMN_218639
Cxxc6	11.26261	36.35355	3.2	ILMN_209685
Glipr1	33.70589	108.0936	3.2	ILMN_217063
9030013K10Rik	41.52975	132.8418	3.2	ILMN_207071
2900016J09Rik	11.79499	37.67979	3.2	ILMN_202127
Ssh3	31.16196	98.67797	3.2	ILMN_214682
Nicn1	186.5011	588.7534	3.2	ILMN_209291
Npr3	28.55296	89.73986	3.1	ILMN_206823
Tnpo2	21.43903	67.03648	3.1	ILMN_189809
Sox9	318.6924	992.889	3.1	ILMN_261259
LOC100044736	25.83275	80.30547	3.1	ILMN_198762
Snf1lk2	14.53984	45.19496	3.1	ILMN_211380
Ebf1	8.861051	27.51272	3.1	ILMN_191261
Xylb	9.435558	29.28097	3.1	ILMN_187649
Gas6	75.23415	232.8612	3.1	ILMN_217682
Rgs4	41.13184	127.2028	3.1	ILMN_209130
Irak3	19.7384	60.54045	3.1	ILMN_222700
Ndufs6	894.5782	2738.651	3.1	ILMN_197367
Mfap2	40.4502	122.9058	3.0	ILMN_209358
Dlx2	33.78724	101.2805	3.0	ILMN_239308
Paip1	40.83055	121.2337	3.0	ILMN_222633
Naprt1	20.79523	61.41785	3.0	ILMN_217024
G630019A19Rik	21.8438	64.49814	3.0	ILMN_207404
Col8a1	176.0184	519.6152	3.0	ILMN_194118
Eng	21.82842	64.33217	2.9	ILMN_193131
Rgma	33.76603	98.832	2.9	ILMN_212164
Adra2a	11.91381	34.825	2.9	ILMN_190996
Rarb	46.7001	136.2455	2.9	ILMN_195865
9630009N10Rik	27.24883	79.26035	2.9	ILMN_205004
Hoxd9	9.394702	27.19967	2.9	ILMN_223278
Prkar1b	27.11168	78.22446	2.9	ILMN_212314
LOC668631	123.3728	353.5466	2.9	ILMN_204532
Mgst2	121.8022	348.8108	2.9	ILMN_211231
Galnt11	210.2867	602.1761	2.9	ILMN_218803
Hoxa4	80.69389	229.8069	2.8	ILMN_222859
Fosb	13.25606	37.65029	2.8	ILMN_192812
Socs2	128.9201	364.4951	2.8	ILMN_212590
Uchl1	81.98615	231.1058	2.8	ILMN_190011
Itgb7	57.49219	162.0092	2.8	ILMN_216930
Hyou1	41.41079	116.6316	2.8	ILMN_215592
Pbx1	59.45525	167.3667	2.8	ILMN_206010
Npr2	79.7575	224.3988	2.8	ILMN_218454
9530050F08Rik	13.24299	37.24749	2.8	ILMN_203683

Mustn1	39.40601	109.074	2.8	ILMN_215350
Leprel2	94.27553	260.4487	2.8	ILMN_214406
Ifit3	18.29643	50.42126	2.8	ILMN_214979
Tbc1d9	4.971491	13.65993	2.7	ILMN_213352
D230007K08Rik	22.13857	60.65035	2.7	ILMN_196597
LOC384344	19.6266	53.612	2.7	ILMN_200542
B430305P08Rik	93.89063	253.1376	2.7	ILMN_196622
6230427J02Rik	28.16925	75.88644	2.7	ILMN_234696
5930418K15Rik	314.4155	844.5696	2.7	ILMN_184478
Rap1gap	6.885259	18.45432	2.7	ILMN_201583
Sema4f	5.613576	15.01651	2.7	ILMN_235029
scl0015365.1_6	35.12809	93.76427	2.7	ILMN_196128
Spry3	29.26513	77.88371	2.7	ILMN_253053
Klf2	144.532	383.1184	2.7	ILMN_210280
D14Ertd500e	12.69582	33.64906	2.7	ILMN_211056
Hyi	163.4672	432.8415	2.6	ILMN_218094
Car12	135.7726	358.8157	2.6	ILMN_233734
H2-DMa	18.12323	47.67715	2.6	ILMN_224639
1700038P13Rik	49.55024	130.3037	2.6	ILMN_196167
5730437N04Rik	640.7289	1675.853	2.6	ILMN_216779
Plau	304.2441	795.5578	2.6	ILMN_212726
Fetub	11.11528	28.87808	2.6	ILMN_223266
6330444G18Rik	51.74722	134.1408	2.6	ILMN_206187
2600001B17Rik	167.2629	432.6639	2.6	ILMN_193322
Hemk1	157.5149	404.6358	2.6	ILMN_216005
Ramp2	22.14312	56.55408	2.6	ILMN_215597
Pde1a	40.85752	104.273	2.6	ILMN_202741
Dlx1	174.8717	444.464	2.5	ILMN_210579
D14Ertd449e	178.3184	451.9613	2.5	ILMN_184960
D930046M13Rik	13.07281	33.09542	2.5	ILMN_191508
Npal1	27.16578	68.61846	2.5	ILMN_222626
Zfhx2	50.62733	127.5678	2.5	ILMN_197499
Icam1	12.53678	31.56981	2.5	ILMN_221313
BC034090	26.3971	66.3407	2.5	ILMN_211522
Crabp1	71.50598	179.4944	2.5	ILMN_218661
Gstk1	15.46747	38.75112	2.5	ILMN_213974
Thbd	96.77809	241.8538	2.5	ILMN_223542
Elovl4	133.1449	330.7346	2.5	ILMN_223323
Gng2	36.52029	90.54585	2.5	ILMN_222219
Ndrg2	5.3967	13.31886	2.5	ILMN_214204
Ogn	245.1254	602.8912	2.5	ILMN_220162
LOC433955	898.34	2203.803	2.5	ILMN_200277
BC066028	42.17014	103.2892	2.4	ILMN_225896
Arhgef3	120.2757	294.5768	2.4	ILMN_214179
St3gal6	108.6663	265.5305	2.4	ILMN_187032
BC017612	73.74236	179.9507	2.4	ILMN_212906
E030025L23Rik	13.01494	31.72122	2.4	ILMN_206889
Oaf	24.63974	59.98651	2.4	ILMN_222394
Cyp4v3	23.30914	56.7099	2.4	ILMN_222960
Armcx6	16.58153	40.2494	2.4	ILMN_249108
Arhgdig	100.9985	244.8166	2.4	ILMN_220986
Fbxl16	14.27613	34.5073	2.4	ILMN_214895
Atg9b	46.27785	111.6558	2.4	ILMN_249271
Ripk3	11.63392	28.01658	2.4	ILMN_222495
Zdhhc3	60.33446	145.1022	2.4	ILMN_184449
4930513E20Rik	28.38401	67.70391	2.4	ILMN_220780
Abhd14a	61.3405	145.8852	2.4	ILMN_252670
LOC100045019	78.12558	184.8972	2.4	ILMN_195453
Gdf1	80.18184	189.3237	2.4	ILMN_214494
Tnc	248.6555	585.5164	2.4	ILMN_185280
1700012H17Rik	138.6384	326.3513	2.4	ILMN_188924

Sdc2	637.5183	1500.019	2.4	ILMN_211837
Galm	28.16032	66.15855	2.3	ILMN_217755
Smpdl3a	18.20438	42.67522	2.3	ILMN_209738
LOC382050	15.72106	36.77679	2.3	ILMN_201166
Fbxo2	60.16906	140.4765	2.3	ILMN_212537
3110001A05Rik	95.39771	221.7077	2.3	ILMN_202154
Abhd1	42.46539	98.59059	2.3	ILMN_216579
Lrig3	586.4861	1361.595	2.3	ILMN_220870
Rab27a	41.87727	96.96421	2.3	ILMN_211354
1700029G01Rik	29.96721	69.35058	2.3	ILMN_221884
LOC100044683	54.02951	125.0313	2.3	ILMN_184844
Car13	274.284	631.3511	2.3	ILMN_219489
Cbr3	330.6771	755.8992	2.3	ILMN_210755
Ccl5	26.22889	59.6003	2.3	ILMN_212715
Ank	790.2815	1795.097	2.3	ILMN_210121
B130051A04Rik	11.42203	25.8662	2.3	ILMN_206009
3202002H23Rik	47.31332	107.0751	2.3	ILMN_206575
Tdrkh	77.96048	176.2869	2.3	ILMN_219057
LOC100047427	39.78378	89.79955	2.3	ILMN_218228
D630004B07Rik	12.16109	27.44156	2.3	ILMN_205082
C130045I22Rik	42.02686	94.58928	2.3	ILMN_192205
BC028528	55.06386	123.8616	2.2	ILMN_215178
Mxd4	45.16882	101.6006	2.2	ILMN_223182
LOC100047264	20.02926	44.98558	2.2	ILMN_194238
Phf19	93.99767	210.584	2.2	ILMN_210062
Inha	12.66026	28.27518	2.2	ILMN_210068
6530401I16Rik	10.33348	23.05144	2.2	ILMN_207046
9530062G19Rik	11.81631	26.35743	2.2	ILMN_203679
Myl9	214.8979	479.1199	2.2	ILMN_214832
H2-D1	147.1148	327.5954	2.2	ILMN_196741
LOC223672	20.99286	46.71019	2.2	ILMN_216469
LOC333744	46.88903	104.3195	2.2	ILMN_198234
C230080E09Rik	18.11628	40.21581	2.2	ILMN_207443
LOC100047082	9.676561	21.47702	2.2	ILMN_220663
Adora2b	127.0244	281.7839	2.2	ILMN_222330
Gfer	1210.959	2686.2	2.2	ILMN_217434
A130021C14Rik	13.72513	30.34787	2.2	ILMN_204839
Plcl2	29.05411	64.10551	2.2	ILMN_187236
Slc26a6	42.79084	94.14424	2.2	ILMN_216189
Arid3a	51.33039	112.6834	2.2	ILMN_217809
Ccdc80	222.9575	489.375	2.2	ILMN_186636
A730040I05Rik	39.32325	85.9285	2.2	ILMN_205840
D9Wsu20e	140.6607	305.8981	2.2	ILMN_186315
LOC675899	1233.75	2683.046	2.2	ILMN_222997
Csgalnact1	7.597406	16.51687	2.2	ILMN_204811
4930546H06Rik	17.08434	37.1077	2.2	ILMN_220063
Dgat2	109.8296	237.8404	2.2	ILMN_211789
LOC100046744	179.8641	388.7363	2.2	ILMN_217084
Ccdc92	127.7451	275.0344	2.2	ILMN_185720
Wnt10a	67.33057	144.8134	2.2	ILMN_189304
Gsta4	56.02418	120.4891	2.2	ILMN_216097
C030027H14Rik	90.48592	194.2234	2.1	ILMN_195794
A930026I22Rik	35.41186	75.57338	2.1	ILMN_188757
Pvrl3	100.2871	213.5925	2.1	ILMN_190689
Cxx1a	21.36272	45.49716	2.1	ILMN_226958
scl0002069.1_48	40.70155	86.62703	2.1	ILMN_187328
Prkg1	48.01555	101.9183	2.1	ILMN_198467
E130010O04Rik	7.984832	16.89067	2.1	ILMN_205411
Arhgap28	12.20895	25.7356	2.1	ILMN_208859
Ngfb	130.6328	274.4039	2.1	ILMN_215493
Gadd45gip1	8.402637	17.63996	2.1	ILMN_213437

D5Wsu178e	342.1687	717.3511	2.1	ILMN_215613
Loxl4	38.04776	79.75448	2.1	ILMN_212724
Vgll3	182.9765	382.511	2.1	ILMN_189588
LOC245892	337.4286	705.3447	2.1	ILMN_198932
Mtrf1	36.89901	77.09094	2.1	ILMN_199371
Dnm3os	67.99136	141.8683	2.1	ILMN_191797
5730410E15Rik	19.07853	39.80537	2.1	ILMN_222288
Tnnc1	86.37764	179.9406	2.1	ILMN_194064
6430526O11Rik	10.06878	20.95958	2.1	ILMN_219257
Atp10d	54.81351	114.0972	2.1	ILMN_210812
2610034E13Rik	10.92902	22.72379	2.1	ILMN_185860
Prrx1	89.05343	185.1128	2.1	ILMN_252720
Camkk1	34.80649	72.3487	2.1	ILMN_216860
1810049O03Rik	15.70355	32.61583	2.1	ILMN_222449
6330404C01Rik	142.181	294.8482	2.1	ILMN_201857
Sepm	58.18999	120.1749	2.1	ILMN_219316
Mlkl	115.1545	237.7563	2.1	ILMN_194213
Slc25a42	25.68548	52.77723	2.1	ILMN_239705
4930481F22Rik	6.107635	12.51629	2.0	ILMN_196062
Dock5	8.200668	16.79367	2.0	ILMN_189346
A930005H10Rik	118.0345	241.512	2.0	ILMN_194765
Atp10a	114.1721	233.4588	2.0	ILMN_220676
Ptafr	6.031271	12.32963	2.0	ILMN_201584
G630038I01Rik	11.37447	23.24307	2.0	ILMN_207278
4930406P12Rik	15.60108	31.84407	2.0	ILMN_202345
E330016A19Rik	401.6247	814.797	2.0	ILMN_207189
Dbp	151.3446	306.3671	2.0	ILMN_211471
A530079D03Rik	13.44146	27.009	2.0	ILMN_204718
Dusp1	514.8914	1034.062	2.0	ILMN_212121
Thsd7a	22.42929	45.03861	2.0	ILMN_200561
Lrrc8	17.54469	35.20151	2.0	ILMN_222070
C230004H03Rik	26.1807	52.47404	2.0	ILMN_205949
2210411K11Rik	386.9212	775.1173	2.0	ILMN_221521
Fhl1	953.698	1907.582	2.0	ILMN_245416
Adamts4	246.6198	123.3096	-2.0	ILMN_214117
Smarcd3	199.396	99.65034	-2.0	ILMN_215257
E030017M08Rik	24.562	12.25078	-2.0	ILMN_205683
Slc7a2	17.1946	8.56276	-2.0	ILMN_191970
Bcl3	133.9487	66.69832	-2.0	ILMN_222447
2610019N19Rik	51.93466	25.8384	-2.0	ILMN_196256
7030401E22Rik	90.36264	44.93195	-2.0	ILMN_206630
C230048N02Rik	28.02898	13.9333	-2.0	ILMN_206433
Fbln1	592.9532	294.4361	-2.0	ILMN_188080
Map4k3	1280.42	635.2591	-2.0	ILMN_215807
LOC100046963	31.92482	15.83429	-2.0	ILMN_194113
Ccdc57	31.05096	15.38112	-2.0	ILMN_223765
LOC100038894	822.5236	407.257	-2.0	ILMN_196719
Usp21	307.9538	152.0818	-2.0	ILMN_193302
Tmem119	661.7032	326.0667	-2.0	ILMN_219443
Slc25a29	26.05035	12.82704	-2.0	ILMN_209032
Sec16b	90.3781	44.43853	-2.0	ILMN_211545
Eif4e	53.90336	26.47637	-2.0	ILMN_210783
Dcamkl1	46.82301	22.92971	-2.0	ILMN_204046
Gramd4	63.78131	31.20813	-2.0	ILMN_208963
Fkbp11	1473.775	720.4687	-2.0	ILMN_210599
Ccng2	129.3723	63.00871	-2.1	ILMN_218789
Atpbd1b	836.7203	406.5232	-2.1	ILMN_214811
A130039H09Rik	74.8169	36.30077	-2.1	ILMN_204052
Pex19	516.311	249.1191	-2.1	ILMN_209858
9530064J02	3120.009	1505.023	-2.1	ILMN_213325
Pyroxd1	353.7421	170.2215	-2.1	ILMN_223861

scl0002507.1_236	1151.946	552.4913	-2.1	ILMN_193522
A530023B05Rik	25.63301	12.28669	-2.1	ILMN_204882
1810043M20Rik	38.80471	18.59451	-2.1	ILMN_201764
Snapc2	25.83515	12.36306	-2.1	ILMN_217540
Rai3	389.9095	186.2403	-2.1	ILMN_223863
Slc9a3r1	1531.305	730.4385	-2.1	ILMN_212760
Dusp28	74.64001	35.35544	-2.1	ILMN_194314
Clstn1	87.25806	41.21202	-2.1	ILMN_187140
Wnt5b	26.95174	12.65898	-2.1	ILMN_191856
Hr	55.96965	26.25727	-2.1	ILMN_218559
Notum	20.99659	9.824379	-2.1	ILMN_194267
Mical2	119.61	55.68366	-2.1	ILMN_221807
E530017G24Rik	27.42919	12.75811	-2.1	ILMN_206095
Prl2c2	389.9386	180.6848	-2.2	ILMN_201515
Wwc1	194.3236	89.99316	-2.2	ILMN_229795
Ppargc1b	74.8576	34.58081	-2.2	ILMN_223929
7330410H16Rik	82.26654	37.94758	-2.2	ILMN_189986
C230055K05Rik	69.46963	32.02577	-2.2	ILMN_229958
Kctd1	254.0371	117.0958	-2.2	ILMN_222835
Mnt	28.73423	13.19558	-2.2	ILMN_230843
Swap70	474.0649	217.4255	-2.2	ILMN_206014
Ppl	74.84332	34.3091	-2.2	ILMN_194079
Tnfaip3	177.1985	81.0732	-2.2	ILMN_187343
Cyb5	1183.336	540.9611	-2.2	ILMN_224117
Aqp5	171.8436	78.49655	-2.2	ILMN_217154
Spsb4	108.8479	49.59155	-2.2	ILMN_212461
Ndg2	141.2222	64.31139	-2.2	ILMN_222397
9630007E23Rik	93.64247	42.37317	-2.2	ILMN_204282
Sobp	50.12623	22.60043	-2.2	ILMN_192584
6330414G02Rik	506.1029	228.1077	-2.2	ILMN_193331
BC064033	108.7913	48.99652	-2.2	ILMN_213776
Prl2c3	832.8683	373.8724	-2.2	ILMN_201546
Hmgcr	68.96626	30.95217	-2.2	ILMN_220196
Tinagl	576.2538	258.4109	-2.2	ILMN_223197
Ybx3	2141.43	956.9924	-2.2	ILMN_203309
B230334I05Rik	130.5165	58.22475	-2.2	ILMN_204518
Dhrs7	296.9062	132.3158	-2.2	ILMN_213586
Serpina3g	932.218	415.089	-2.2	ILMN_220729
9630009A08Rik	29.33263	13.01438	-2.3	ILMN_204210
Man1c1	41.40658	18.35849	-2.3	ILMN_211041
D230017C05Rik	98.99113	43.86392	-2.3	ILMN_205625
5430435G22Rik	157.4943	69.73066	-2.3	ILMN_223803
Trim37	31.42459	13.90569	-2.3	ILMN_194081
Spata7	18.21226	8.057454	-2.3	ILMN_210281
Tmc4	69.59656	30.79008	-2.3	ILMN_186472
4632415K11Rik	95.61253	42.19706	-2.3	ILMN_214381
3110021A11Rik	72.25201	31.84795	-2.3	ILMN_189769
Prl2c4	575.0294	253.4433	-2.3	ILMN_199380
Pdlim2	295.9025	130.3903	-2.3	ILMN_211617
Lgals2	26.40362	11.62195	-2.3	ILMN_213844
2010011I20Rik	40.96349	17.9658	-2.3	ILMN_218925
Mapkapk3	211.5999	92.72988	-2.3	ILMN_224132
Coq3	44.89563	19.56421	-2.3	ILMN_219194
Socs3	1053.615	458.2059	-2.3	ILMN_211651
Capns2	33.79014	14.6572	-2.3	ILMN_185539
2610044O15Rik	48.87528	21.14706	-2.3	ILMN_217556
D17H6S56E-5	150.9485	65.00861	-2.3	ILMN_196579
Gata6	701.6214	301.7988	-2.3	ILMN_209474
2410051C13Rik	64.07466	27.42427	-2.3	ILMN_202217
Edn2	29.23084	12.44106	-2.3	ILMN_220576
Hoxb6	500.6657	212.9976	-2.4	ILMN_216818

Ifi35	195.0465	82.46265	-2.4	ILMN_212338
Vegfa	402.8679	169.9742	-2.4	ILMN_224512
5830411K21Rik	83.54332	35.24244	-2.4	ILMN_186422
Ctsc	296.5959	125.0957	-2.4	ILMN_214875
Agtrap	112.0101	47.22925	-2.4	ILMN_208935
Runx1	77.90806	32.84678	-2.4	ILMN_192279
Lbh	1467.945	618.3139	-2.4	ILMN_215742
LOC100047645	29.81422	12.51626	-2.4	ILMN_196644
Zfp593	25.38135	10.5977	-2.4	ILMN_190772
1190002H23Rik	504.4489	210.4599	-2.4	ILMN_209224
Heg1	43.15241	17.97368	-2.4	ILMN_235037
Itpk1	215.241	89.33578	-2.4	ILMN_220572
2200002D01Rik	160.0353	66.30216	-2.4	ILMN_184461
Ralgps2	133.6022	55.19207	-2.4	ILMN_238709
Manba	256.1089	105.703	-2.4	ILMN_219952
Oxr1	107.5049	44.09302	-2.4	ILMN_213669
EG632802	67.85917	27.80243	-2.4	ILMN_196874
scl0002775.1_4	95.30786	38.99197	-2.4	ILMN_186200
Mkx	200.8846	82.07319	-2.4	ILMN_219001
C130023A14Rik	150.7899	61.24025	-2.5	ILMN_195997
Areg	223.8034	90.41296	-2.5	ILMN_217903
Slc7a11	65.43837	26.43497	-2.5	ILMN_205646
Asgr1	293.6649	118.2188	-2.5	ILMN_210744
Ddr1	551.9943	221.9004	-2.5	ILMN_219438
Nek6	35.23449	14.15863	-2.5	ILMN_217629
Bnc1	156.6407	62.88671	-2.5	ILMN_223056
Omp	21.2088	8.514393	-2.5	ILMN_184266
Leprel1	42.21786	16.90505	-2.5	ILMN_184387
A730017D01Rik	204.0831	81.47868	-2.5	ILMN_187956
AI646023	172.9942	68.7208	-2.5	ILMN_210462
Aste1	138.311	54.71539	-2.5	ILMN_212823
Paqr7	135.6327	53.37876	-2.5	ILMN_216225
En1	151.9833	59.76232	-2.5	ILMN_261544
Ttc23	17.04704	6.698587	-2.5	ILMN_215477
Nt5e	51.74896	20.27041	-2.6	ILMN_213360
Tcf7	51.18078	20.02527	-2.6	ILMN_215616
LOC100046616	43.50359	17.00677	-2.6	ILMN_217154
Map3k6	99.42783	38.86625	-2.6	ILMN_233949
Kif26b	79.52576	31.08009	-2.6	ILMN_211222
Ccdc109b	52.51559	20.51011	-2.6	ILMN_209057
Sult2b1	17.55362	6.797498	-2.6	ILMN_217684
Osgin1	133.4348	51.14267	-2.6	ILMN_214233
Pld1	268.6967	102.951	-2.6	ILMN_209394
Wisp2	881.7182	337.7096	-2.6	ILMN_192308
Arfl4	33.94349	13.00025	-2.6	ILMN_212066
Gmpr	383.1645	146.1067	-2.6	ILMN_210136
F11r	166.077	63.19294	-2.6	ILMN_219759
Niban	52.58195	19.93658	-2.6	ILMN_191711
Apobec3	89.72931	33.95382	-2.6	ILMN_218267
Rab3b	23.92311	9.017391	-2.7	ILMN_193861
4930404F20Rik	45.86389	17.25426	-2.7	ILMN_202239
2210408F11Rik	380.9002	143.28	-2.7	ILMN_191967
Khk	433.3749	162.9312	-2.7	ILMN_218651
Amigo2	69.83593	26.17855	-2.7	ILMN_193798
6330509M05Rik	54.60982	20.45814	-2.7	ILMN_187875
LOC100048436	491.1138	183.2299	-2.7	ILMN_220948
Plvap	42.7189	15.89207	-2.7	ILMN_216938
1600002H07Rik	44.77868	16.64469	-2.7	ILMN_214668
Ckb	2722.313	1011.862	-2.7	ILMN_193661
LOC676640	245.3346	91.02916	-2.7	ILMN_184785
Adamts7	37.53836	13.90093	-2.7	ILMN_212242

A130038J17Rik	96.7347	35.78524	-2.7	ILMN_195241
6720422M22Rik	327.0119	120.5923	-2.7	ILMN_195056
Tmem106a	52.39495	19.20371	-2.7	ILMN_220104
Dcp1b	48.90105	17.88965	-2.7	ILMN_237203
Sulf2	377.8372	137.9805	-2.7	ILMN_191187
Them4	34.08429	12.44703	-2.7	ILMN_218844
Cdh3	97.0977	34.89893	-2.8	ILMN_228163
Pla1a	19.4513	6.989331	-2.8	ILMN_236152
Lonrf3	45.82985	16.43044	-2.8	ILMN_222854
Bckdhb	65.47566	23.4725	-2.8	ILMN_215543
Rbpms	121.3852	43.4986	-2.8	ILMN_190688
2810003C17Rik	2016.479	721.0569	-2.8	ILMN_221837
Zfp52	58.79827	21.00913	-2.8	ILMN_196018
Wnt6	73.08739	25.97512	-2.8	ILMN_185059
A330021E22Rik	49.08216	17.42338	-2.8	ILMN_212595
Crabp2	542.6907	192.396	-2.8	ILMN_223221
Glrx1	672.1139	238.1391	-2.8	ILMN_224227
LOC229810	288.1544	101.7644	-2.8	ILMN_197643
2810408P10Rik	88.46747	31.04037	-2.9	ILMN_209085
LOC100044177	553.576	192.6378	-2.9	ILMN_185055
Plekha6	70.95644	24.63108	-2.9	ILMN_211786
Gpr146	25.43071	8.813067	-2.9	ILMN_258519
Dusp4	137.0223	47.35087	-2.9	ILMN_188216
Rdh10	36.47513	12.51385	-2.9	ILMN_221783
B230343A10Rik	284.5547	96.58751	-2.9	ILMN_190249
Csdc2	39.26384	13.32289	-2.9	ILMN_221533
Tspan12	36.40434	12.34344	-2.9	ILMN_190691
Adrb2	147.343	49.75607	-3.0	ILMN_217517
Rin3	186.6556	62.76797	-3.0	ILMN_215946
Ptgs1	49.90517	16.76317	-3.0	ILMN_213920
Chd7	33.62745	11.20729	-3.0	ILMN_215704
Hoxb9	40.16719	13.34236	-3.0	ILMN_224060
Sphk1	244.0817	80.86204	-3.0	ILMN_221190
C030014I23Rik	36.79984	12.16953	-3.0	ILMN_196060
Pcdh21	198.6133	65.67675	-3.0	ILMN_216921
2410146L05Rik	33.01647	10.83134	-3.0	ILMN_214200
Sfrp1	583.1172	190.8084	-3.1	ILMN_210306
Atp7a	55.86957	18.26951	-3.1	ILMN_219616
3300005D01Rik	91.3115	29.8568	-3.1	ILMN_210819
Vdr	208.1276	67.82571	-3.1	ILMN_187559
A730054J21Rik	857.3325	277.4537	-3.1	ILMN_187889
Epb4.1l4a	415.7589	133.9888	-3.1	ILMN_194065
Gdpd3	49.96845	16.10321	-3.1	ILMN_217647
Spp1	3260.93	1048.658	-3.1	ILMN_218026
Eme2	48.53425	15.54953	-3.1	ILMN_216685
LOC674427	220.9294	70.69267	-3.1	ILMN_207028
Sv2a	178.8575	57.06665	-3.1	ILMN_218994
9030611O19Rik	193.7802	61.0526	-3.2	ILMN_222981
Sox11	719.3066	225.6198	-3.2	ILMN_214025
Aldh3b1	180.7815	56.64622	-3.2	ILMN_214234
Cacna2d1	127.6848	39.73623	-3.2	ILMN_204931
Anpep	163.685	50.72736	-3.2	ILMN_208800
Mgat3	228.6955	69.61978	-3.3	ILMN_215046
LOC100047651	2867.046	866.6475	-3.3	ILMN_187579
Mylk	413.6839	124.0413	-3.3	ILMN_190234
Clmn	40.52944	12.10616	-3.3	ILMN_208976
Atpbd3	61.59446	18.35896	-3.4	ILMN_222461
Nme5	59.95346	17.84005	-3.4	ILMN_184565
Nudt6	164.3049	48.66026	-3.4	ILMN_210970
St6galnac2	822.9109	242.3379	-3.4	ILMN_201528
Rapgefl1	117.5419	34.44387	-3.4	ILMN_212357

Camk2n1	301.7679	87.75297	-3.4	ILMN_254160
Cdon	344.0735	99.95779	-3.4	ILMN_203979
Ndrl	916.241	264.6656	-3.5	ILMN_214267
Dlx3	56.78472	16.26912	-3.5	ILMN_216829
Gsdmdc1	204.9081	58.67153	-3.5	ILMN_222389
Arrdc3	209.5952	59.96544	-3.5	ILMN_216006
Slpi	60.99187	17.08165	-3.6	ILMN_219640
Parp3	191.5368	53.56898	-3.6	ILMN_211968
E530016P10Rik	49.73941	13.90828	-3.6	ILMN_206561
Ptpn6	73.65567	20.50577	-3.6	ILMN_222531
Stard6	54.5612	15.12968	-3.6	ILMN_220243
BC032265	32.06116	8.880066	-3.6	ILMN_224038
Prkag2	125.9228	34.82687	-3.6	ILMN_221101
1200016E24Rik	177.6524	48.86094	-3.6	ILMN_191167
Otx1	728.431	200.1859	-3.6	ILMN_217984
4732435K05Rik	80.65807	22.01444	-3.7	ILMN_203523
Megf10	108.0106	28.55141	-3.8	ILMN_198326
Grem1	2915.548	767.2842	-3.8	ILMN_213216
OTTMUSG00000010673	310.0212	81.18607	-3.8	ILMN_228043
Cotl1	1548.189	403.7357	-3.8	ILMN_219070
Raet1b	409.9851	105.9697	-3.9	ILMN_201542
Il11	126.6849	32.71162	-3.9	ILMN_215774
G430005B15Rik	85.06609	21.81089	-3.9	ILMN_207399
Nod1	537.3145	137.3481	-3.9	ILMN_222316
1810015C04Rik	210.8496	53.74314	-3.9	ILMN_222164
Pdgfb	273.1614	69.51267	-3.9	ILMN_211703
Lama5	606.1924	154.05	-3.9	ILMN_213736
Epm2aip1	88.64892	22.38274	-4.0	ILMN_218771
Cd59a	119.8463	30.1365	-4.0	ILMN_214646
Adssl1	164.0222	40.9008	-4.0	ILMN_210075
Podxl	77.34526	19.2795	-4.0	ILMN_210330
2610024M03Rik	601.1432	148.653	-4.0	ILMN_189885
1190007F08Rik	368.3851	90.46323	-4.1	ILMN_232083
Col4a5	455.1432	110.848	-4.1	ILMN_186711
2700017A04Rik	23.9494	5.805688	-4.1	ILMN_221682
Rgs17	673.4044	162.6912	-4.1	ILMN_187084
Trim56	156.6218	36.83594	-4.3	ILMN_187187
Rpl29	2543.357	573.9031	-4.4	ILMN_221019
Hpse	127.0154	28.63137	-4.4	ILMN_216095
Ndrg1	564.7465	125.8694	-4.5	ILMN_223174
Spon2	1450.617	320.2812	-4.5	ILMN_223108
Cyp2s1	72.52116	15.9488	-4.5	ILMN_223026
LOC100047870	172.2172	37.54394	-4.6	ILMN_213638
Ncald	67.49866	14.67383	-4.6	ILMN_214828
Rapgef3	92.61622	19.93019	-4.6	ILMN_213815
Kcnab1	150.6522	31.77479	-4.7	ILMN_211650
Rpp25	263.8009	55.43745	-4.8	ILMN_222749
Parp8	102.1306	21.12031	-4.8	ILMN_209706
Centd3	82.13421	16.65382	-4.9	ILMN_211864
Mgst1	580.4053	113.1717	-5.1	ILMN_216289
Hoxc9	944.6088	182.5156	-5.2	ILMN_223992
Smoc1	64.23833	12.37951	-5.2	ILMN_211655
Cox7a2l	1226.658	234.3773	-5.2	ILMN_223726
8030402P03Rik	434.4914	82.38298	-5.3	ILMN_207113
Il33	82.16874	15.34694	-5.4	ILMN_213496
Cxadr	574.4451	105.16	-5.5	ILMN_227636
Scube3	225.8754	40.23154	-5.6	ILMN_249967
Hs3st1	188.8531	33.59804	-5.6	ILMN_209971
Cxcl15	131.9494	23.34883	-5.7	ILMN_213274
Odz4	207.1419	35.81351	-5.8	ILMN_194814
Cryab	294.8189	50.81403	-5.8	ILMN_217805

Adh7	518.7874	86.74728	-6.0	ILMN_211984
Galntl4	171.061	27.80354	-6.2	ILMN_221507
Msln	234.4364	37.39478	-6.3	ILMN_211608
Kcnh2	117.686	18.31039	-6.4	ILMN_213233
Zbtb7c	108.8779	16.83062	-6.5	ILMN_212080
1110059G02Rik	120.9926	18.07338	-6.7	ILMN_190837
Aim1	915.7984	136.606	-6.7	ILMN_208931
Baiap2l1	206.5662	29.47513	-7.0	ILMN_215954
Dio3	128.2648	18.16983	-7.1	ILMN_223923
2810423A18Rik	1375.56	182.5468	-7.5	ILMN_186667
Igfbp2	158.5099	20.07263	-7.9	ILMN_222264
Cdh26	261.5187	32.75253	-8.0	ILMN_219271
Axud1	269.8358	33.10211	-8.2	ILMN_213373
Etl4	96.33768	11.37184	-8.5	ILMN_192240
Wnt4	128.3642	14.94104	-8.6	ILMN_195108
Mmp10	109.9946	12.6962	-8.7	ILMN_211834
Smpd3	297.9785	33.24168	-9.0	ILMN_220010
Mmp13	52.81427	5.718719	-9.2	ILMN_210384
Kcnmb4	82.58699	8.850835	-9.3	ILMN_213638
Hmha1	257.1252	23.94753	-10.7	ILMN_217580
Igfbp4	4144.354	351.4114	-11.8	ILMN_210225
Elfn1	178.3773	13.59004	-13.1	ILMN_210519
Lars2	231.2172	16.64084	-13.9	ILMN_221567
Uap1	1112.49	72.77073	-15.3	ILMN_194014
Cp	180.4986	10.43172	-17.3	ILMN_203167
D1Ertd471e	562.78	31.32249	-18.0	ILMN_187044
Mrps27	276.6797	15.28519	-18.1	ILMN_208650
1110018J23Rik	606.5533	30.06133	-20.2	ILMN_213005
Sprr2k	266.1004	13.09267	-20.3	ILMN_258141
Igsf4a	1275.006	44.73253	-28.5	ILMN_194372
Ptgis	1588.106	44.51039	-35.7	ILMN_220660
LOC383308	2168.66	33.4696	-64.8	ILMN_200537

Supplementary Table 5.2 Transcripts significantly mis-regulated in Non-transformed (nko) Mbd4$^{-/-}$ cells. Shown are all transcripts mis-regulated >3-fold in ntMbd4$^{-/-}$ cells (relative to wildtype cells) and their average signals on the BeadChips. Positive number in the fold change column indicates upregulation in knockout cells; negative number means downregulation.

Gene Name	Wildtype_AVG_Signal	Knockout_AVG_Signal	Fold Change	SEARCH_KEY
C1qtnf3	10.34066	3475.413	336.1	ILMN_209151
Fhl1	46.74125	5953.684	127.4	ILMN_245416
Ptgis	34.74647	4075.586	117.3	ILMN_220660
Igfbp2	80.54948	8795.384	109.2	ILMN_222264
Col8a1	53.97717	1955.796	36.2	ILMN_194118
1110032E23Rik	65.62099	2363.884	36.0	ILMN_187154
Pappa	32.79083	781.3681	23.8	ILMN_212814
Inhba	77.75533	1510.099	19.4	ILMN_209192
Sdpr	240.0267	4613.933	19.2	ILMN_217783
Hebp1	24.20243	464.4074	19.2	ILMN_218714
Tiam1	102.4417	1920.403	18.7	ILMN_221605
Cdkn1a	65.35168	1128.013	17.3	ILMN_209664
Rai3	35.60147	562.067	15.8	ILMN_223863
Klf4	9.76973	154.0405	15.8	ILMN_209216
Tspan2	21.27343	328.587	15.4	ILMN_186229

Mcam	86.23905	1289.41	15.0	ILMN_211310
Gprc5a	59.87022	892.6079	14.9	ILMN_239682
5033414K04Rik	67.74311	961.5439	14.2	ILMN_210073
9130213B05Rik	95.78693	1242.969	13.0	ILMN_216760
Eng	32.9481	396.9045	12.0	ILMN_193131
Adrb2	30.69232	360.1751	11.7	ILMN_217517
Npy1r	29.65161	318.9252	10.8	ILMN_211592
Chrd	33.40907	357.5493	10.7	ILMN_223518
Mgll	48.96025	517.6797	10.6	ILMN_209140
Dennd2a	69.15369	678.6058	9.8	ILMN_218566
scl0003073.1_164	37.92841	354.0112	9.3	ILMN_186746
Limch1	163.2591	1512.866	9.3	ILMN_212627
Zcchc5	33.84047	310.5942	9.2	ILMN_187122
Tnc	186.2198	1706.848	9.2	ILMN_185280
Cryab	254.9966	2314.04	9.1	ILMN_217805
Rspo3	126.0854	1138.259	9.0	ILMN_221501
Pkp1	17.66537	146.3172	8.3	ILMN_223293
Sulf1	184.4747	1490.345	8.1	ILMN_193870
Tspan12	39.25039	312.4616	8.0	ILMN_190691
2200002D01Rik	60.42404	468.2123	7.7	ILMN_184461
Fyb	13.86576	107.092	7.7	ILMN_222124
Snurf	84.76928	651.1382	7.7	ILMN_221855
Gdf15	146.5044	1119.177	7.6	ILMN_221419
Uchl1	34.47517	263.1922	7.6	ILMN_190011
Thbs1	23.68723	179.5712	7.6	ILMN_215409
Rasl11b	142.6339	1058.608	7.4	ILMN_200046
LOC100048332	232.5773	1680.558	7.2	ILMN_193758
D730042P09Rik	16.70375	116.1017	7.0	ILMN_202674
Dusp4	36.55875	245.4398	6.7	ILMN_188216
Ccng1	287.1877	1900.458	6.6	ILMN_193770
2810003C17Rik	278.8297	1797.36	6.4	ILMN_221837
Epb4.1l4b	67.34761	429.5949	6.4	ILMN_224014
Prelp	349.159	2159.948	6.2	ILMN_221743
Osbpl3	278.8793	1725.145	6.2	ILMN_214782
Dcxr	158.7735	958.6804	6.0	ILMN_223082
Dgkg	155.7137	935.0571	6.0	ILMN_188868
Il17d	21.43921	128.6117	6.0	ILMN_248374
2210015K02Rik	30.57849	177.7166	5.8	ILMN_202282
LOC227393	27.35723	157.4897	5.8	ILMN_197084
3110082D06Rik	47.2532	267.1554	5.7	ILMN_195121
LOC100043986	39.69152	221.5186	5.6	ILMN_219654
5031436O03Rik	169.1642	934.8342	5.5	ILMN_189777
Cryl1	85.18723	467.1452	5.5	ILMN_217939
Arsj	28.17718	153.8121	5.5	ILMN_217748
Mical2	69.74276	372.9203	5.3	ILMN_221807
Fgfrl1	166.3217	880.064	5.3	ILMN_211983
Tspan18	28.42294	149.8678	5.3	ILMN_210130
6230400G14Rik	81.13354	423.0014	5.2	ILMN_193477
Ccl17	17.206	89.6376	5.2	ILMN_220052
Anxa4	119.6601	618.8853	5.2	ILMN_223968
Raet1b	77.2205	397.822	5.2	ILMN_201542
Upk3b	11.51574	59.31396	5.2	ILMN_188123
Slc19a2	86.88435	441.0793	5.1	ILMN_223225
Myom1	86.27552	437.7882	5.1	ILMN_241505
Nod1	105.4804	531.3724	5.0	ILMN_222316
LOC381860	16.40346	82.07124	5.0	ILMN_200791
Efnb2	20.39202	101.7812	5.0	ILMN_219995
Slc7a2	14.31631	71.37851	5.0	ILMN_191970
Sort1	63.70035	316.7707	5.0	ILMN_212031
Bmper	51.59573	256.5308	5.0	ILMN_223608
Zmat3	211.7886	1042.415	4.9	ILMN_184413

Pkia	236.4293	1155.107	4.9	ILMN_216692
Nupr1	1175.275	5636.806	4.8	ILMN_221920
BC064033	39.56429	189.1685	4.8	ILMN_213776
LOC100048169	24.88584	117.1807	4.7	ILMN_185828
LOC100040657	17.47009	81.3749	4.7	ILMN_199303
Papss2	172.6139	795.1919	4.6	ILMN_184209
Klf2	225.2548	1035.931	4.6	ILMN_210280
Igfbp5	17.48514	80.13703	4.6	ILMN_190841
Nuak1	962.6436	4407.984	4.6	ILMN_234029
Tes	426.914	1934.907	4.5	ILMN_216487
E230020D15Rik	82.30367	370.571	4.5	ILMN_184584
Ndn	438.6167	1946.532	4.4	ILMN_212068
Lims2	159.8899	708.5914	4.4	ILMN_221324
Llgl2	236.2876	1040.838	4.4	ILMN_223065
Acta1	125.4867	551.6194	4.4	ILMN_221671
Dner	171.9009	751.3356	4.4	ILMN_219944
Tead4	28.05703	122.1389	4.4	ILMN_192395
Creb5	17.32844	75.17724	4.3	ILMN_217830
Crim1	15.10852	65.38578	4.3	ILMN_195784
AW146242	216.4223	919.8475	4.3	ILMN_221426
Bdnf	98.04885	415.0121	4.2	ILMN_216070
4933427D14Rik	19.6879	83.27302	4.2	ILMN_220964
Ctgf	3573.182	15098.79	4.2	ILMN_223979
LOC100043822	39.72166	167.7831	4.2	ILMN_188198
E130308A19Rik	11.58619	48.87088	4.2	ILMN_207120
Tbxa2r	28.6958	120.9408	4.2	ILMN_219522
BC022687	51.94145	218.6448	4.2	ILMN_212921
Flrt3	127.707	535.0834	4.2	ILMN_210479
Picalm	118.2471	493.0108	4.2	ILMN_209744
Snx30	51.28943	213.0861	4.2	ILMN_185019
AI646023	235.9839	979.0677	4.1	ILMN_210462
Ngfrap1	3500.679	14478.04	4.1	ILMN_211730
Col4a5	175.5711	725.7273	4.1	ILMN_186711
9930004G02Rik	28.35655	117.1744	4.1	ILMN_204385
Nxn	164.6491	679.738	4.1	ILMN_190874
Trp53inp1	120.6186	497.4815	4.1	ILMN_194391
Slit2	84.78352	349.5049	4.1	ILMN_186672
D4Bwg0951e	816.8505	3354.776	4.1	ILMN_214279
2900027G03Rik	21.30311	87.19423	4.1	ILMN_214582
Hbegf	255.0348	1043.609	4.1	ILMN_218639
1110006E14Rik	37.05984	151.6167	4.1	ILMN_188726
Whrn	48.62869	196.9041	4.0	ILMN_229647
Raet1c	12.95481	52.16194	4.0	ILMN_201541
4632401N01Rik	92.09451	369.8586	4.0	ILMN_202709
3110013H01Rik	60.08421	240.5877	4.0	ILMN_215880
1110036O03Rik	645.123	2555.411	4.0	ILMN_214424
Sephs2	31.15468	122.7231	3.9	ILMN_209348
D0H4S114	1071.905	4200.065	3.9	ILMN_217188
scl000171.1_4	38.25821	149.1599	3.9	ILMN_195535
Phlda3	472.6763	1831.998	3.9	ILMN_212624
Gcap27	3259.445	12627.13	3.9	ILMN_186123
Exoc4	63.48079	243.6419	3.8	ILMN_209872
Glipr1	661.6081	2512.467	3.8	ILMN_217063
Gsk3b	64.48595	244.3067	3.8	ILMN_191246
6030400N17Rik	27.90663	105.5988	3.8	ILMN_203014
Mtap6	32.27213	121.9607	3.8	ILMN_206927
scl0003781.1_22	86.64725	326.4056	3.8	ILMN_196280
1700038P13Rik	89.6554	336.2522	3.8	ILMN_196167
Fas	99.92044	373.4239	3.7	ILMN_188640
Creg1	25.9222	94.55373	3.6	ILMN_211269
2510022D24Rik	26.56856	96.71023	3.6	ILMN_185937

Pcgf3	48.53363	176.631	3.6	ILMN_190917
A930010I20Rik	17.69053	64.05283	3.6	ILMN_211923
Garnl3	92.3194	331.1686	3.6	ILMN_217308
E230025E14Rik	18.1512	65.03798	3.6	ILMN_195795
C920027I18Rik	62.2155	222.883	3.6	ILMN_191962
A130028J20Rik	49.53832	176.6129	3.6	ILMN_195564
Gmfg	8.721038	31.07448	3.6	ILMN_217353
Vgll3	299.5761	1067.052	3.6	ILMN_189588
Magi1	11.85661	41.76398	3.5	ILMN_195326
Tpd52	115.0247	404.6506	3.5	ILMN_188233
Gas6	896.6461	3142.542	3.5	ILMN_217682
C330023J09Rik	55.25688	193.5969	3.5	ILMN_207088
Pmepa1	57.57077	200.8496	3.5	ILMN_186309
2900026A02Rik	134.7628	470.1495	3.5	ILMN_187380
Rspo2	532.2729	1839.279	3.5	ILMN_215554
A230106J09Rik	25.20563	85.5309	3.4	ILMN_202683
Tceal8	1826.084	6179.885	3.4	ILMN_223469
Dpysl3	189.3699	638.8594	3.4	ILMN_218073
Bax	184.0618	615.2999	3.3	ILMN_211762
Ostm1	726.2676	2424.744	3.3	ILMN_212196
Chrnb1	54.42599	180.3282	3.3	ILMN_217651
Mdm2	823.8137	2724.952	3.3	ILMN_220130
Tmed8	12.65355	41.70699	3.3	ILMN_261943
Trim11	31.97846	104.9074	3.3	ILMN_186471
Hes1	153.5354	503.4648	3.3	ILMN_210390
Ncam1	71.89225	235.3249	3.3	ILMN_184370
C230060E24	39.44452	128.8912	3.3	ILMN_219479
Npr3	221.5217	723.5316	3.3	ILMN_206823
9530081N05Rik	81.99737	266.8964	3.3	ILMN_205593
Sox9	850.7388	2765.737	3.3	ILMN_261259
Lmod1	142.6916	463.5367	3.2	ILMN_210366
Rgs4	228.0729	737.4663	3.2	ILMN_209130
Pkp2	83.91284	270.9414	3.2	ILMN_213446
Ctsh	213.0318	686.6871	3.2	ILMN_241653
D830014B20Rik	116.8066	375.7643	3.2	ILMN_206756
F630107L03Rik	46.90493	150.4308	3.2	ILMN_207418
LOC100044736	28.32229	90.70721	3.2	ILMN_198762
Psrc1	21.84832	69.66103	3.2	ILMN_218163
Cap1	330.7706	1054.483	3.2	ILMN_214505
Endod1	163.0493	514.6422	3.2	ILMN_221277
Csdc2	20.70249	64.75287	3.1	ILMN_221533
A230021I18Rik	31.0249	96.96757	3.1	ILMN_204870
Serpinb1a	127.0207	395.6677	3.1	ILMN_218572
Dysf	130.7893	406.6433	3.1	ILMN_184715
AW123240	57.65371	179.0297	3.1	ILMN_195216
Reck	80.15815	248.8711	3.1	ILMN_223820
4933421H10Rik	111.2033	344.9431	3.1	ILMN_221469
Plscr3	23.54509	72.93366	3.1	ILMN_218814
Gpr23	270.6616	835.2703	3.1	ILMN_223959
Rusc2	398.6543	1228.811	3.1	ILMN_214316
BB146404	98.15958	301.51	3.1	ILMN_184649
Ptrh1	215.5212	654.2502	3.0	ILMN_208796
B930083E23Rik	71.27987	215.3564	3.0	ILMN_205609
1500015O10Rik	44.01895	132.7568	3.0	ILMN_222808
B230312A22Rik	161.0088	484.5546	3.0	ILMN_210245
Evi1	15.02089	45.08184	3.0	ILMN_214877
D230005E09Rik	88.69312	266.1036	3.0	ILMN_204935
H19	26.42925	79.26706	3.0	ILMN_219881
Psmb5	721.7653	240.4372	-3.0	ILMN_211192
Hic2	201.6055	67.10759	-3.0	ILMN_194097
9830002I17Rik	50.6334	16.83948	-3.0	ILMN_221228

B830017H08Rik	61.35537	20.37897	-3.0	ILMN_203494
Matn2	95.43671	31.69371	-3.0	ILMN_220046
Mkx	306.1007	101.5028	-3.0	ILMN_219001
BC021381	279.3847	92.3955	-3.0	ILMN_223208
Slco1a5	76.1471	25.17623	-3.0	ILMN_221283
4833409N03Rik	506.1417	167.3305	-3.0	ILMN_195376
A930005H10Rik	389.5107	128.7569	-3.0	ILMN_194765
4930445K14Rik	451.5424	149.0859	-3.0	ILMN_189952
9330188B09Rik	96.36654	31.80994	-3.0	ILMN_204370
Sphk1	261.9549	85.98535	-3.0	ILMN_221190
Rnf213	1798.633	590.383	-3.0	ILMN_213503
Sesn3	41.1235	13.49722	-3.0	ILMN_184128
Zfp41	607.1316	198.9374	-3.1	ILMN_189448
Tmem55b	1350.659	441.6316	-3.1	ILMN_221957
Hoxa4	268.5713	87.64173	-3.1	ILMN_222859
Aprin	1006.914	328.5021	-3.1	ILMN_216438
AA409316	338.2637	110.2374	-3.1	ILMN_221994
B430201A12Rik	274.7596	89.42561	-3.1	ILMN_192167
8030467N07Rik	105.9939	34.47438	-3.1	ILMN_207056
AI413782	37.92953	12.33626	-3.1	ILMN_214512
9130004J05Rik	189.9596	61.74068	-3.1	ILMN_193459
1110031I02Rik	110.9074	36.0303	-3.1	ILMN_222011
4831403C07Rik	180.3356	58.55266	-3.1	ILMN_188701
4930550C14Rik	145.0882	47.0752	-3.1	ILMN_213960
Fgd6	242.289	78.60153	-3.1	ILMN_224288
Pcyt2	328.9428	106.6649	-3.1	ILMN_209615
scl0002624.1_576	95.85807	31.07387	-3.1	ILMN_184713
Sdc3	2006.668	648.0548	-3.1	ILMN_219906
Elfn1	250.0585	80.70979	-3.1	ILMN_210519
2310040C09Rik	84.39429	27.21946	-3.1	ILMN_186638
Glt8d2	165.4777	53.35516	-3.1	ILMN_203977
Zfp820	82.39539	26.5363	-3.1	ILMN_204952
6230425C21Rik	500.8293	161.2033	-3.1	ILMN_202500
Sqrdl	118.4931	38.1079	-3.1	ILMN_223305
Runx1t1	229.0862	73.61584	-3.1	ILMN_210856
Slc23a3	62.98718	20.23257	-3.1	ILMN_219791
8030481K01Rik	134.5237	43.18512	-3.1	ILMN_207013
Clec11a	347.7948	111.5976	-3.1	ILMN_189820
Samd4	114.8198	36.83779	-3.1	ILMN_221466
AI256775	166.9932	53.46572	-3.1	ILMN_185879
Eno3	2266.861	724.3914	-3.1	ILMN_222979
Phf13	355.161	113.2705	-3.1	ILMN_214014
2310005L22Rik	1581.353	503.5748	-3.1	ILMN_192733
Mccc2	95.26058	30.23473	-3.2	ILMN_219661
LOC100044683	202.0879	63.97835	-3.2	ILMN_184844
1110049F12Rik	434.6578	137.4501	-3.2	ILMN_212471
Npff	81.31773	25.70894	-3.2	ILMN_216935
Tnrc6b	36.78185	11.62456	-3.2	ILMN_204914
Glrx	4128.174	1303.683	-3.2	ILMN_238781
Scnn1a	49.29381	15.55668	-3.2	ILMN_220993
Errfi1	2069.818	653.1835	-3.2	ILMN_190042
Snta1	243.7582	76.92226	-3.2	ILMN_221335
B330002M01Rik	56.29587	17.75134	-3.2	ILMN_203490
Cdkn2c	489.1868	154.0961	-3.2	ILMN_213040
Nanos1	162.3268	50.7773	-3.2	ILMN_218879
Steap1	362.5601	113.2835	-3.2	ILMN_210027
1700021C14Rik	220.056	68.62109	-3.2	ILMN_221114
C230009C22Rik	55.37031	17.25932	-3.2	ILMN_207368
Mme	301.0988	93.7514	-3.2	ILMN_206056
BC038925	96.06147	29.88059	-3.2	ILMN_213504
C030003A18Rik	302.6373	93.59549	-3.2	ILMN_203064

Pogz	29.03939	8.978961	-3.2	ILMN_229041
scl0002007.1_97	175.2955	54.08537	-3.2	ILMN_191549
ENSMUSG00000074822	36.00861	11.10246	-3.2	ILMN_206176
Pdk1	111.1515	34.23186	-3.2	ILMN_214993
P2rx5	164.3631	50.51679	-3.3	ILMN_221894
Dlk2	427.0789	131.1975	-3.3	ILMN_218738
2310014F06Rik	72.20573	22.17144	-3.3	ILMN_192976
Bicc1	2458.86	754.8414	-3.3	ILMN_215610
Col6a1	7706.899	2365.454	-3.3	ILMN_208779
5430414B19Rik	53.47944	16.41045	-3.3	ILMN_196359
Sipa1l2	166.2525	50.96694	-3.3	ILMN_223616
Tlcd1	380.4151	116.6059	-3.3	ILMN_212241
Zhx2	193.4011	59.15037	-3.3	ILMN_195577
Slc24a3	1891.465	578.3742	-3.3	ILMN_220855
Prrx1	126.8333	38.78068	-3.3	ILMN_252720
Klhl17	834.7981	254.4461	-3.3	ILMN_209660
Thbs2	3580.359	1088.005	-3.3	ILMN_213253
Otx1	710.0976	215.7306	-3.3	ILMN_217984
C130057N11Rik	179.8773	54.62923	-3.3	ILMN_190624
2700063P19Rik	226.4756	68.57703	-3.3	ILMN_202082
B230312E02Rik	180.925	54.76945	-3.3	ILMN_206789
Itgb4	58.42581	17.64483	-3.3	ILMN_213200
Hlx	313.1851	94.50337	-3.3	ILMN_213507
Hist1h2ae	57.19297	17.21672	-3.3	ILMN_196748
6430407L02Rik	282.5469	84.88407	-3.3	ILMN_203250
1700025G04Rik	269.3204	80.72125	-3.3	ILMN_216343
Zfp316	103.5618	31.0323	-3.3	ILMN_184408
Hexdc	303.8002	90.96548	-3.3	ILMN_241904
Map3k1	285.8687	85.50298	-3.3	ILMN_211294
Esco2	359.0276	107.1411	-3.4	ILMN_208759
9330177P20Rik	358.2766	106.8308	-3.4	ILMN_195505
Tapbp	71.2259	21.22718	-3.4	ILMN_213284
Pltp	105.1166	31.30003	-3.4	ILMN_208776
Spry3	120.7332	35.88885	-3.4	ILMN_253053
C130040J23Rik	277.2737	82.07254	-3.4	ILMN_192554
D930007N19Rik	75.41334	22.29677	-3.4	ILMN_195777
Tnfrsf4	62.56816	18.49527	-3.4	ILMN_187722
Scube3	334.1714	98.56088	-3.4	ILMN_249967
Aak1	41.24004	12.14643	-3.4	ILMN_254339
Rpl22	2970.654	874.1182	-3.4	ILMN_215417
Dok3	126.6829	37.20255	-3.4	ILMN_217284
Polg2	119.163	34.95186	-3.4	ILMN_228682
LOC100039623	49.54964	14.53228	-3.4	ILMN_196792
Letmd1	740.5508	217.1092	-3.4	ILMN_211469
Sbsn	67.79766	19.62374	-3.5	ILMN_216457
Tuba4a	50.09013	14.47289	-3.5	ILMN_195615
Cxcr7	109.6363	31.66273	-3.5	ILMN_217785
Clip4	154.9312	44.6416	-3.5	ILMN_193898
2610029K11Rik	71.20475	20.46325	-3.5	ILMN_189412
Rab28	4764.253	1369.139	-3.5	ILMN_196462
Arhgdig	487.7996	139.3873	-3.5	ILMN_220986
EG433229	6663.32	1903.48	-3.5	ILMN_215511
Suv420h1	240.3018	68.48883	-3.5	ILMN_186777
Mboat1	1030.688	292.6428	-3.5	ILMN_209663
Pdxp	893.0643	253.4111	-3.5	ILMN_219636
Ramp2	302.8783	85.91326	-3.5	ILMN_215597
Lnx2	396.9727	112.3743	-3.5	ILMN_209763
Rnf207	38.87628	11.0043	-3.5	ILMN_196652
Ankrd28	389.287	109.8226	-3.5	ILMN_187988
Sfrp1	2983.888	839.7314	-3.6	ILMN_210306
Tmem53	416.5225	117.0727	-3.6	ILMN_215216

4922503N01Rik	150.7895	42.37249	-3.6	ILMN_210268
5430426F23Rik	64.01309	17.91711	-3.6	ILMN_205677
Samd1	83.52781	23.26606	-3.6	ILMN_211130
BC016495	88.67927	24.68978	-3.6	ILMN_216548
D17H6S56E-5	78.48188	21.82955	-3.6	ILMN_196579
Shd	44.71228	12.38441	-3.6	ILMN_236512
scl0003799.1_2	1751.105	483.5699	-3.6	ILMN_186828
Tmem179b	709.0638	195.7274	-3.6	ILMN_215148
Ccdc9	39.60387	10.90828	-3.6	ILMN_216266
Csf2ra	72.7596	20.0078	-3.6	ILMN_211976
Mreg	51.29582	14.06098	-3.6	ILMN_253687
Grem1	383.5378	105.0664	-3.7	ILMN_213216
LOC333331	80.17326	21.9444	-3.7	ILMN_215850
Itga2b	96.77187	26.40188	-3.7	ILMN_221217
C130002N06	73.85911	20.10927	-3.7	ILMN_186859
Lrrc17	286.5811	78.02413	-3.7	ILMN_194980
Col17a1	52.21601	14.13935	-3.7	ILMN_221356
Slc43a1	167.063	45.18572	-3.7	ILMN_193227
Ccl27	151.1404	40.78555	-3.7	ILMN_216432
Tmem119	438.8438	117.817	-3.7	ILMN_219443
Tasp1	54.61119	14.65575	-3.7	ILMN_219832
Emp2	1888.514	505.6946	-3.7	ILMN_209674
2810021B07Rik	211.2101	56.50565	-3.7	ILMN_210854
Il1rn	662.9731	176.613	-3.8	ILMN_217380
Slc27a6	52.67462	14.02646	-3.8	ILMN_214291
Hsf4	61.30737	16.32078	-3.8	ILMN_222242
Zfp825	49.34717	13.02571	-3.8	ILMN_220879
Map3k6	131.917	34.81593	-3.8	ILMN_233949
Nme3	316.575	83.4649	-3.8	ILMN_222087
Tlr2	375.3636	98.61851	-3.8	ILMN_221308
B430305P08Rik	212.8502	55.8018	-3.8	ILMN_196622
Acss2	175.2679	45.83664	-3.8	ILMN_216644
D730003I15Rik	46.2605	12.06949	-3.8	ILMN_195420
Epdr1	3530.29	919.6412	-3.8	ILMN_214840
Map4k2	202.8405	52.68924	-3.8	ILMN_188020
Samd10	107.2754	27.76465	-3.9	ILMN_215170
4932408C11Rik	216.0285	55.87397	-3.9	ILMN_203468
Mapk8ip1	127.7168	33.00172	-3.9	ILMN_223186
LOC226017	954.3906	246.1977	-3.9	ILMN_198408
Rgs17	550.531	141.7143	-3.9	ILMN_187084
Tap2	94.12442	24.12951	-3.9	ILMN_217715
LOC433943	255.2281	65.26745	-3.9	ILMN_198112
C230043G09Rik	915.9519	233.9937	-3.9	ILMN_205959
D930042A21Rik	69.27411	17.68757	-3.9	ILMN_206876
Zfp817	71.65195	18.29189	-3.9	ILMN_243422
Ppm2c	117.0272	29.82523	-3.9	ILMN_261727
0610040B10Rik	104.2708	26.44995	-3.9	ILMN_215673
D230048P18Rik	110.7031	28.06521	-3.9	ILMN_206055
Cd248	1559.326	395.1964	-3.9	ILMN_216345
2410008K03Rik	259.9987	65.70718	-4.0	ILMN_223117
Itpk1	88.74877	22.33282	-4.0	ILMN_220572
E130014J05Rik	71.25277	17.89008	-4.0	ILMN_189509
Stmn2	65.94501	16.52906	-4.0	ILMN_210469
Rpgrip1l	189.9234	47.56342	-4.0	ILMN_221013
Glrx1	2839.261	709.8499	-4.0	ILMN_224227
AI747699	96.80217	24.0925	-4.0	ILMN_258552
2310001H17Rik	102.5508	25.52005	-4.0	ILMN_213451
Hps6	192.1075	47.74895	-4.0	ILMN_223360
Syn1	252.9124	62.79257	-4.0	ILMN_222720
Recql4	199.7906	49.37078	-4.0	ILMN_211352
Cfi	59.32181	14.64562	-4.1	ILMN_212927

8030402P03Rik	703.345	173.1073	-4.1	ILMN_207113
C1qtnf6	128.8834	31.67092	-4.1	ILMN_221628
C920004C08Rik	220.3085	54.055	-4.1	ILMN_190891
Sobp	145.8139	35.56796	-4.1	ILMN_192584
Mustn1	424.0905	103.2088	-4.1	ILMN_215350
Ifi27	69.35394	16.84204	-4.1	ILMN_184190
Tle2	1280.361	310.7886	-4.1	ILMN_184889
Sv2a	38.79099	9.414949	-4.1	ILMN_218994
Angptl6	53.96983	13.05633	-4.1	ILMN_214036
Pcdhb22	333.9556	80.32935	-4.2	ILMN_213754
Ltbp4	562.2674	135.1103	-4.2	ILMN_208840
Sepp1	164.1141	39.32566	-4.2	ILMN_217349
Cml1	63.8477	15.29464	-4.2	ILMN_221685
Ptp4a3	136.8153	32.74582	-4.2	ILMN_213418
D15Bwg0759e	303.4806	72.23535	-4.2	ILMN_195654
1810062O18Rik	80.02505	19.02348	-4.2	ILMN_192867
D930015E06Rik	357.1513	84.12836	-4.2	ILMN_208662
S100a4	492.1362	115.6567	-4.3	ILMN_221174
A430106B04Rik	106.5604	24.95858	-4.3	ILMN_203048
EG638695	118.4819	27.4859	-4.3	ILMN_257853
Nt5dc2	3195.979	736.9973	-4.3	ILMN_215517
Acadm	6587.83	1506.211	-4.4	ILMN_227083
4933436C20Rik	164.0848	37.40826	-4.4	ILMN_210320
Lgals3bp	2344.197	533.7465	-4.4	ILMN_209135
1200016E24Rik	92.13522	20.97804	-4.4	ILMN_191167
LOC223672	88.5661	20.02322	-4.4	ILMN_216469
2610307O08Rik	196.201	44.13772	-4.4	ILMN_221480
Ift80	121.3162	27.2909	-4.4	ILMN_220633
1700027N10Rik	133.2828	29.94068	-4.5	ILMN_242897
Bcl3	176.5321	39.59875	-4.5	ILMN_222447
Fbln1	591.2425	132.433	-4.5	ILMN_188080
LOC380844	215.6401	48.17661	-4.5	ILMN_197188
Trpc2	100.2522	22.37511	-4.5	ILMN_192626
Hoxb7	1917.568	427.6057	-4.5	ILMN_211288
1810062G17Rik	198.0838	43.70291	-4.5	ILMN_208903
Sult2b1	51.21893	11.29793	-4.5	ILMN_217684
LOC100038894	585.8642	128.6927	-4.6	ILMN_196719
Irx3	1654.412	363.0829	-4.6	ILMN_224225
Slc2a6	127.9485	28.07224	-4.6	ILMN_211726
Slpi	292.1718	64.00404	-4.6	ILMN_219640
9030425E11Rik	190.2837	41.6839	-4.6	ILMN_189630
5430411C19Rik	88.68565	19.18629	-4.6	ILMN_193294
Cd9	3428.769	734.9481	-4.7	ILMN_220691
Spon2	1440.58	307.9263	-4.7	ILMN_223108
5730409K12Rik	94.96428	20.10878	-4.7	ILMN_193516
HIP-1	54.45347	11.51115	-4.7	ILMN_206964
Zfpm2	493.7007	103.8086	-4.8	ILMN_195765
A930026I22Rik	140.8073	29.59997	-4.8	ILMN_188757
Cnksr3	152.7074	32.09559	-4.8	ILMN_222298
Sord	331.1094	69.54982	-4.8	ILMN_223021
Supt16h	928.9438	194.0957	-4.8	ILMN_223539
Wnt5b	311.056	64.8243	-4.8	ILMN_191856
1700081H05Rik	109.4135	22.76836	-4.8	ILMN_196392
Gstm1	643.9881	133.9432	-4.8	ILMN_212441
Bcl2l11	300.0225	61.96204	-4.8	ILMN_208904
AB112350	101.7023	20.98958	-4.8	ILMN_211687
Angpt4	74.65689	15.36233	-4.9	ILMN_222129
Cd14	181.7816	37.22334	-4.9	ILMN_221922
Ntn2l	148.5249	30.34522	-4.9	ILMN_210119
Exoc3l	119.4873	24.35279	-4.9	ILMN_218843
Sema5a	42.51192	8.663857	-4.9	ILMN_204130

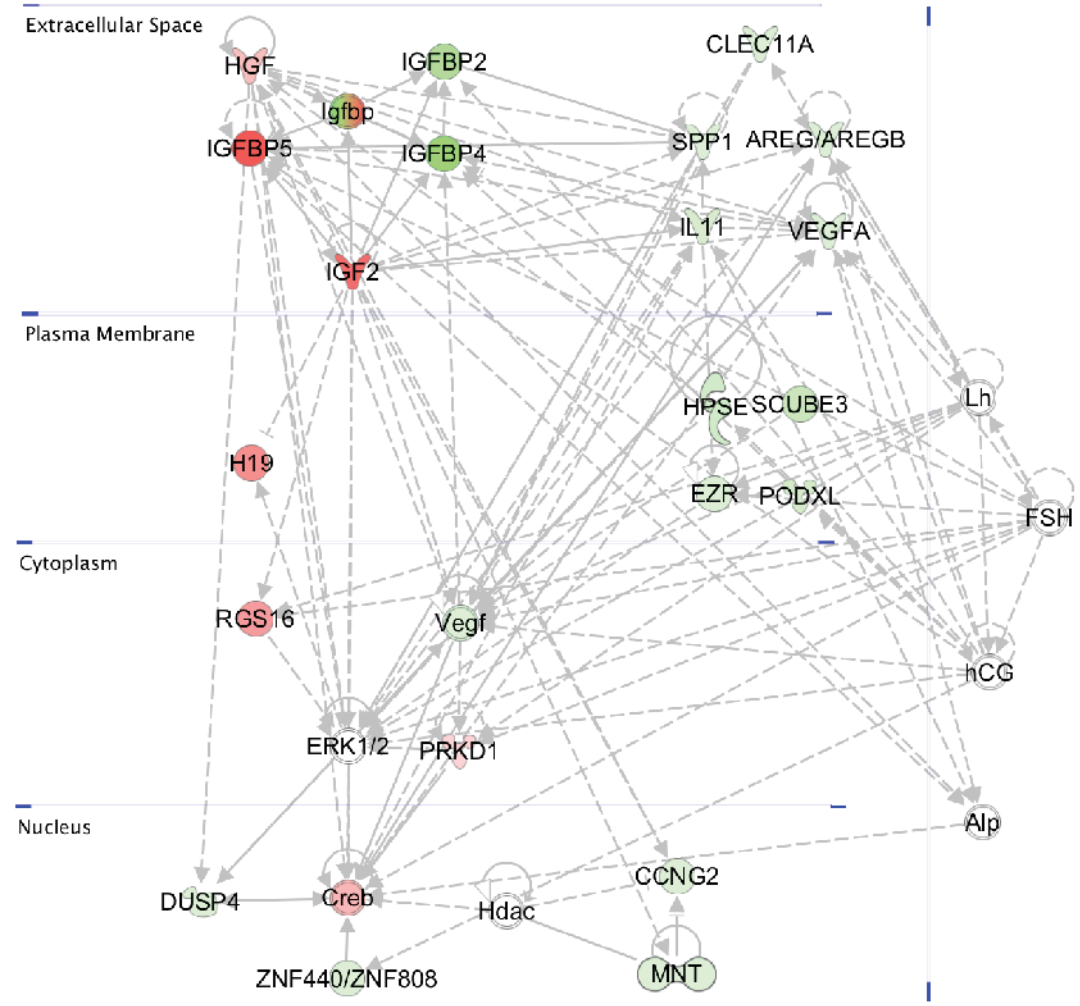
Stk3	49.79597	10.1136	-4.9	ILMN_206738
Mlkl	178.1503	35.9819	-5.0	ILMN_194213
Ssh3	266.1504	53.6825	-5.0	ILMN_214682
4632417K18Rik	918.9	185.114	-5.0	ILMN_220188
Mycl1	222.6399	44.67434	-5.0	ILMN_213909
Ppm1l	162.824	32.45504	-5.0	ILMN_184575
Hoxb5	405.5818	80.26199	-5.1	ILMN_213971
Ankrd47	132.1799	26.12701	-5.1	ILMN_221468
Muc1	57.11061	11.27767	-5.1	ILMN_208888
5031425E22Rik	236.1068	46.07572	-5.1	ILMN_222309
S3-12	1460.003	284.779	-5.1	ILMN_208656
4930572J05Rik	289.8939	55.32214	-5.2	ILMN_219117
2310076O14Rik	154.1566	29.28517	-5.3	ILMN_221888
C230098O21Rik	84.13457	15.91922	-5.3	ILMN_184645
Gch1	49.68622	9.39511	-5.3	ILMN_210506
Tmem35	95.53104	17.98076	-5.3	ILMN_188922
Abi3bp	504.459	94.72248	-5.3	ILMN_212813
Lrp4	564.1052	105.4323	-5.4	ILMN_213961
3110070M22Rik	73.05151	13.64733	-5.4	ILMN_218458
5730593F17Rik	136.6469	25.4998	-5.4	ILMN_223419
Pcsk4	44.60275	8.322991	-5.4	ILMN_212743
Gdf1	314.7358	58.42004	-5.4	ILMN_214494
Slc7a7	44.15223	8.125346	-5.4	ILMN_217989
S100a1	2225.106	407.0105	-5.5	ILMN_214542
OTTMUSG00000010673	293.4787	53.36362	-5.5	ILMN_228043
Vit	633.0846	113.7882	-5.6	ILMN_191294
1500005N04Rik	283.4388	50.68332	-5.6	ILMN_202497
Lbp	466.5725	82.02612	-5.7	ILMN_223901
Il13ra1	823.1803	144.6002	-5.7	ILMN_216227
Fmnl2	1005.092	175.9994	-5.7	ILMN_215164
3110021A11Rik	136.0923	23.69496	-5.7	ILMN_189769
H2-DMa	182.8962	31.70216	-5.8	ILMN_224639
Aldh3b1	636.8285	109.9207	-5.8	ILMN_214234
Cln3	199.2089	34.11724	-5.8	ILMN_220100
Rbm14	99.4691	17.0183	-5.8	ILMN_204298
Cyp4v3	456.1019	77.36152	-5.9	ILMN_222960
Aoc3	774.609	130.758	-5.9	ILMN_203969
A730040I05Rik	174.1803	29.34691	-5.9	ILMN_205840
Cenpp	567.4822	95.17678	-6.0	ILMN_222261
Rgs9bp	73.8065	12.24271	-6.0	ILMN_217635
Ager	183.1248	30.2821	-6.0	ILMN_210862
Map3k8	103.0053	17.02335	-6.1	ILMN_204712
Mad2l2	199.2484	32.72016	-6.1	ILMN_219303
Ppargc1b	104.2315	17.06873	-6.1	ILMN_223929
Mylk	789.2093	128.524	-6.1	ILMN_190234
EG623230	1474.634	239.2042	-6.2	ILMN_186051
Clec2d	484.4626	77.99601	-6.2	ILMN_210243
9430064K01Rik	256.6933	41.17949	-6.2	ILMN_192598
C230066H01Rik	67.49267	10.8124	-6.2	ILMN_206723
Kcnh2	116.1785	18.55562	-6.3	ILMN_213233
Entpd2	123.2936	19.69098	-6.3	ILMN_215332
Egr3	323.6744	51.49716	-6.3	ILMN_195299
Smarca1	1148.928	182.7624	-6.3	ILMN_209637
Dlg7	284.9103	45.2245	-6.3	ILMN_214861
Vcam1	121.1164	19.12282	-6.3	ILMN_202787
6720477C19Rik	540.2004	84.75518	-6.4	ILMN_185354
Ccl7	1192.973	187.1649	-6.4	ILMN_223895
Aldh3a1	544.4032	84.92189	-6.4	ILMN_223471
1190007F08Rik	464.1748	72.40321	-6.4	ILMN_232083
Phf21b	105.7302	16.4435	-6.4	ILMN_214807
LOC381739	1506.708	230.8535	-6.5	ILMN_200488

Rhou	1064.165	163.0326	-6.5	ILMN_215476
Lrrc8d	659.3508	100.7718	-6.5	ILMN_223609
Cda	237.0583	36.12019	-6.6	ILMN_220880
2810423A18Rik	1661.572	252.6184	-6.6	ILMN_186667
Susd2	160.2677	24.12811	-6.6	ILMN_219457
Rasl11a	1649.947	247.4422	-6.7	ILMN_225258
Ccl2	160.9719	24.10948	-6.7	ILMN_211411
Calml4	232.4484	34.68931	-6.7	ILMN_212718
D830050E09Rik	103.7771	15.4724	-6.7	ILMN_205110
Gsdmdc1	233.5528	34.74226	-6.7	ILMN_222389
Cables1	391.4229	57.35958	-6.8	ILMN_216355
Slc1a3	1467.349	214.9821	-6.8	ILMN_213165
Ang	350.5417	51.157	-6.9	ILMN_244375
Psmb8	143.7287	20.80076	-6.9	ILMN_185344
Socs3	1826.607	263.8726	-6.9	ILMN_211651
Slc25a45	490.2873	70.34534	-7.0	ILMN_220462
Pacs1	218.8533	31.26811	-7.0	ILMN_220552
Twist2	1197.49	170.6341	-7.0	ILMN_186022
Dkk3	1015.621	144.5972	-7.0	ILMN_219132
6330414G02Rik	1288.185	182.0019	-7.1	ILMN_193331
Cdkl2	171.6058	24.16768	-7.1	ILMN_211931
Mapk13	103.4825	14.52988	-7.1	ILMN_209546
Mt1	12416.66	1725.369	-7.2	ILMN_209514
Tgfbi	2841.483	394.0177	-7.2	ILMN_192154
Igsf10	291.3417	40.38636	-7.2	ILMN_187617
A230005G17Rik	139.6336	19.16831	-7.3	ILMN_205324
Gstm4	140.2984	19.24629	-7.3	ILMN_220200
LOC638301	199.7959	27.34194	-7.3	ILMN_186725
Tspan17	1277.231	174.4811	-7.3	ILMN_208788
Aatk	126.6958	17.30565	-7.3	ILMN_221321
Tshz3	871.5295	118.3936	-7.4	ILMN_194436
Edg2	253.0988	33.93119	-7.5	ILMN_224105
Parp8	279.3118	37.18604	-7.5	ILMN_209706
Brp17	715.2197	94.842	-7.5	ILMN_213275
H1f0	143.9181	18.5612	-7.8	ILMN_222971
Net1	116.2877	14.72615	-7.9	ILMN_187922
1700029G01Rik	195.8975	24.67467	-7.9	ILMN_221884
Serpina3n	1408.404	176.5648	-8.0	ILMN_222516
Sncg	79.90771	9.994639	-8.0	ILMN_209715
Elovl4	129.0072	16.01648	-8.1	ILMN_223323
Fbxo44	116.8509	14.50542	-8.1	ILMN_223033
Ube1l	341.9855	42.31841	-8.1	ILMN_192098
scl0002507.1_236	967.5753	119.2388	-8.1	ILMN_193522
Scd1	1710.462	209.7351	-8.2	ILMN_218062
2600010E01Rik	103.2605	12.602	-8.2	ILMN_222950
H2-Q7	42.77068	5.219112	-8.2	ILMN_196751
Itih2	409.1718	49.86595	-8.2	ILMN_216672
Htra1	1517.907	183.1698	-8.3	ILMN_222238
D630040I23Rik	255.7581	30.80201	-8.3	ILMN_193224
Col6a2	3466.582	414.9943	-8.4	ILMN_216544
Acy3	178.3488	21.1849	-8.4	ILMN_220901
E130102H24Rik	129.4637	15.03233	-8.6	ILMN_214408
Odz4	124.4277	14.38235	-8.7	ILMN_194814
Gchfr	243.5134	27.7866	-8.8	ILMN_215064
Ndrl	1494.107	168.7862	-8.9	ILMN_214267
Lyz	333.0328	37.26352	-8.9	ILMN_218930
Gpha2	117.2885	12.80067	-9.2	ILMN_211491
Olfm1	127.7846	13.83351	-9.2	ILMN_212106
Ogn	2323.883	250.3238	-9.3	ILMN_220162
Rnf144a	250.3409	26.52234	-9.4	ILMN_217013
LOC100046518	190.1332	20.13372	-9.4	ILMN_221629

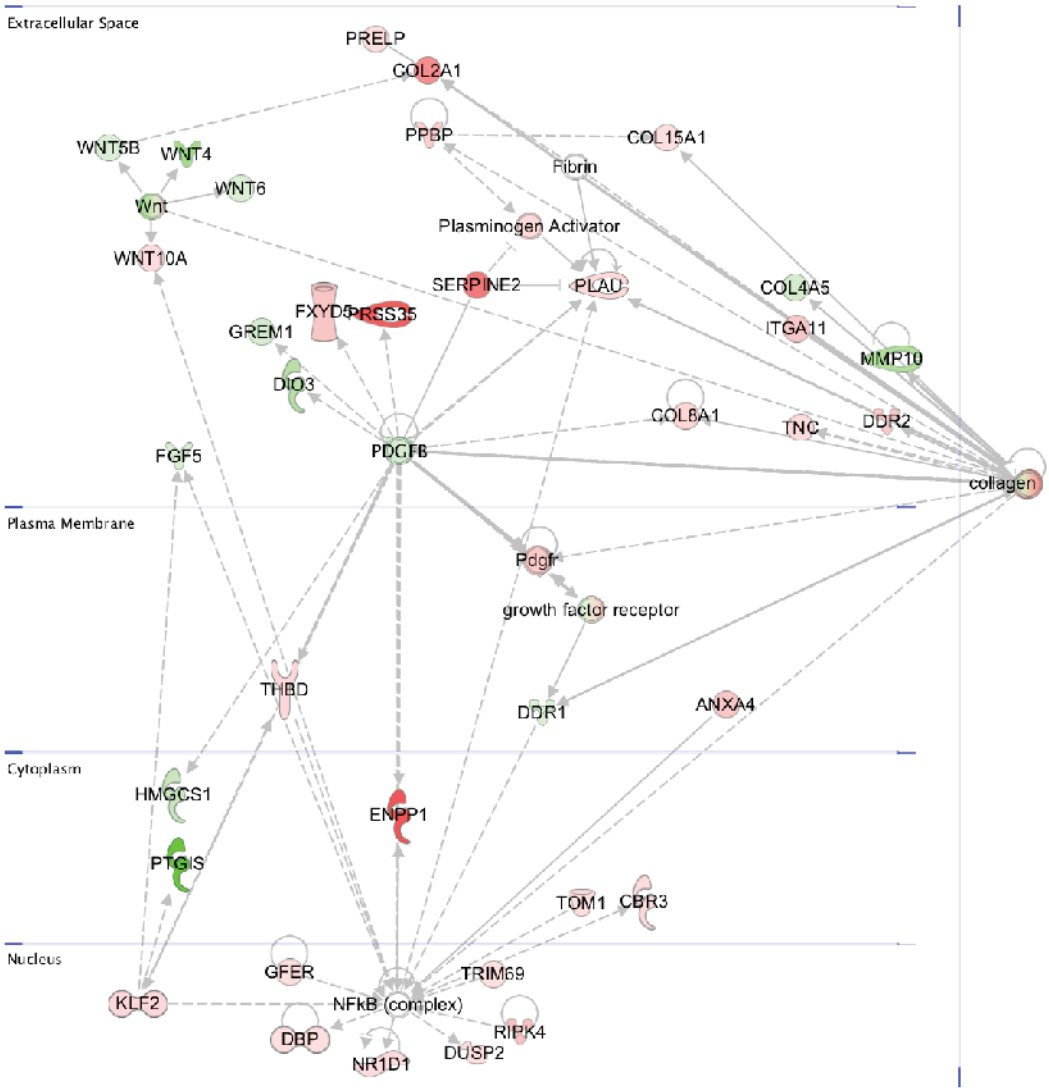
Icam1	203.987	20.73097	-9.8	ILMN_221313
2310039E09Rik	163.7812	16.57272	-9.9	ILMN_211197
Pcp4l1	656.2724	66.22733	-9.9	ILMN_203100
St6galnac2	1688.984	167.5566	-10.1	ILMN_201528
Hapln4	440.497	43.46132	-10.1	ILMN_208828
2810405K02Rik	219.7994	20.9408	-10.5	ILMN_220895
5330431K02Rik	939.2208	88.20355	-10.6	ILMN_219563
C1qdc2	738.8486	68.31789	-10.8	ILMN_214190
Cbx7	170.1508	15.60744	-10.9	ILMN_219110
6720458D17Rik	459.7479	41.88677	-11.0	ILMN_187288
Col14a1	426.9518	38.45116	-11.1	ILMN_203320
Ccno	229.3811	20.56048	-11.2	ILMN_221503
Arhgef3	473.7351	42.0942	-11.3	ILMN_214179
Wnt4	173.0376	14.91518	-11.6	ILMN_195108
LOC268828	203.6957	17.31733	-11.8	ILMN_197760
Plekhg4	420.4597	33.63264	-12.5	ILMN_220732
Gstt1	1954.227	150.2489	-13.0	ILMN_217226
2210016F16Rik	174.9863	13.3278	-13.1	ILMN_218530
Ppp1r3c	1909.399	144.8433	-13.2	ILMN_216078
Scara5	3424.242	257.0929	-13.3	ILMN_222294
Slc27a3	256.2712	18.31193	-14.0	ILMN_222312
Crip2	2657.361	187.6231	-14.2	ILMN_219558
Dio3	353.993	24.98933	-14.2	ILMN_223923
Camk2n1	302.0564	20.83071	-14.5	ILMN_254160
Wif1	460.5247	31.27604	-14.7	ILMN_194714
H2-K1	102.4962	6.769175	-15.1	ILMN_208622
Olfml3	1562.581	99.81375	-15.7	ILMN_219432
scl0002702.1_3805	574.9486	35.83515	-16.0	ILMN_195541
Lgals3	4300.176	264.7988	-16.2	ILMN_222956
Tnnt1	438.6359	26.73763	-16.4	ILMN_189308
Ndrg1	757.7451	46.11279	-16.4	ILMN_223174
Anpep	368.9018	22.38591	-16.5	ILMN_208800
Chi3l1	1127.408	67.26659	-16.8	ILMN_210847
Nnmt	896.3633	53.12735	-16.9	ILMN_201751
Oplah	287.9391	17.04069	-16.9	ILMN_215664
BC028528	438.6346	25.89315	-16.9	ILMN_215178
Smpdl3b	1608.293	88.98137	-18.1	ILMN_210171
Ggt7	347.264	19.15092	-18.1	ILMN_210362
LOC674135	481.4417	24.97606	-19.3	ILMN_220727
Gng13	590.0897	30.56078	-19.3	ILMN_211395
Mmp23	494.2267	24.70391	-20.0	ILMN_215522
Avil	221.6535	11.06891	-20.0	ILMN_222666
Macrod1	567.7158	28.2387	-20.1	ILMN_217502
Pla1a	163.5443	7.981448	-20.5	ILMN_236152
Igf2	1182.897	56.26129	-21.0	ILMN_209642
Egfl7	835.6354	38.93021	-21.5	ILMN_207457
Cd59a	570.6747	26.14909	-21.8	ILMN_214646
Akp2	653.2094	29.12788	-22.4	ILMN_215583
Cd82	989.0947	39.5298	-25.0	ILMN_222273
Pdlim4	3915.495	146.2634	-26.8	ILMN_210491
Hp	1439.596	50.07777	-28.7	ILMN_214144
Smoc1	753.7094	25.70373	-29.3	ILMN_211655
C3	1176.613	39.85423	-29.5	ILMN_223109
Lcn2	1287.872	41.67941	-30.9	ILMN_219674
Mfap2	270.3965	8.334862	-32.4	ILMN_209358
Aqp1	2872.172	85.4954	-33.6	ILMN_216314
Pdgfra	455.4812	13.14808	-34.6	ILMN_212227
LOC100045864	3981.891	111.8958	-35.6	ILMN_196741
Capn6	1363.685	37.21738	-36.6	ILMN_218388
Hoxb6	979.7885	22.98364	-42.6	ILMN_216818
H2-D1	1887.698	44.122	-42.8	ILMN_196741

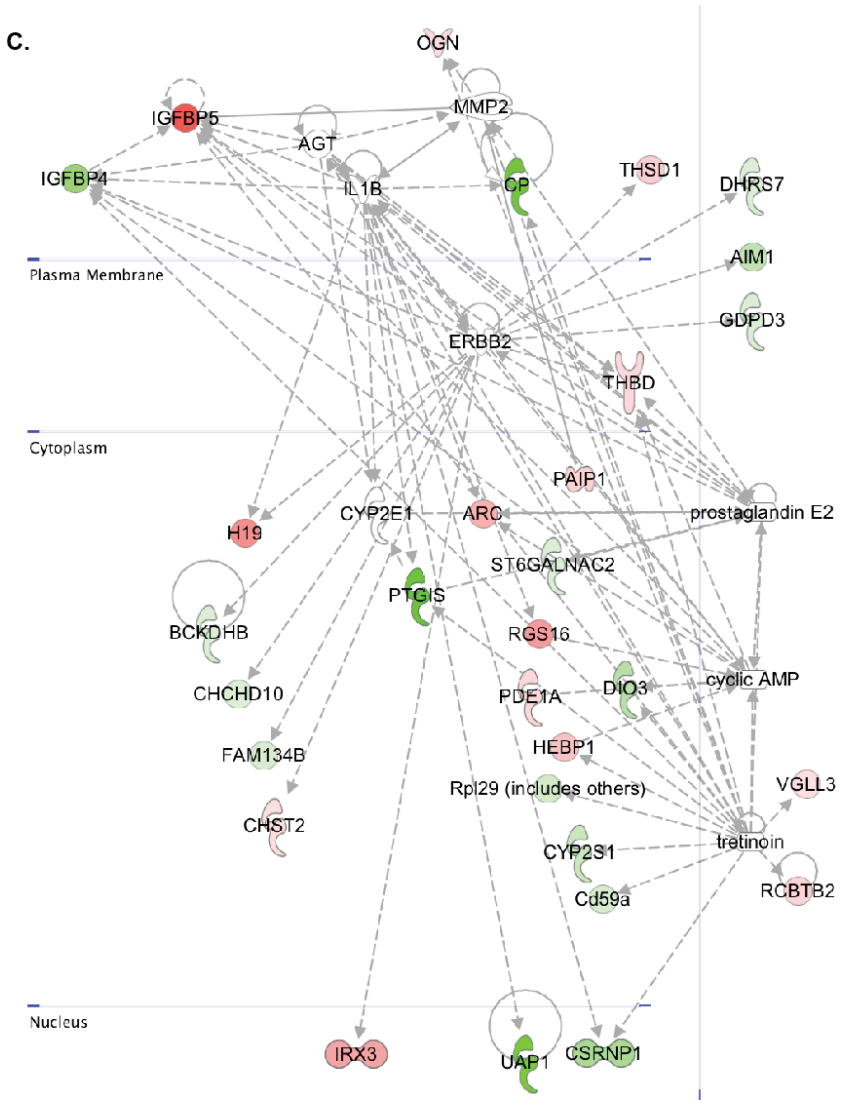
Emilin2	1951.222	42.77086	-45.6	ILMN_250777
Gsta4	528.7911	11.1307	-47.5	ILMN_216097
Fbxo2	1691.625	33.10698	-51.1	ILMN_212537
5430433G21Rik	2827.351	46.44001	-60.9	ILMN_220885
Mgst2	1925.706	31.26836	-61.6	ILMN_211231
Serping1	1084.449	14.94927	-72.5	ILMN_256713
Igfbp4	5869.105	79.47958	-73.8	ILMN_210225
Cpxm1	1485.039	17.4844	-84.9	ILMN_214237
Chst1	1483.803	17.43739	-85.1	ILMN_222161
Pcdh21	1577.516	17.22009	-91.6	ILMN_216921
Mmp13	1277.063	8.618011	-148.2	ILMN_210384

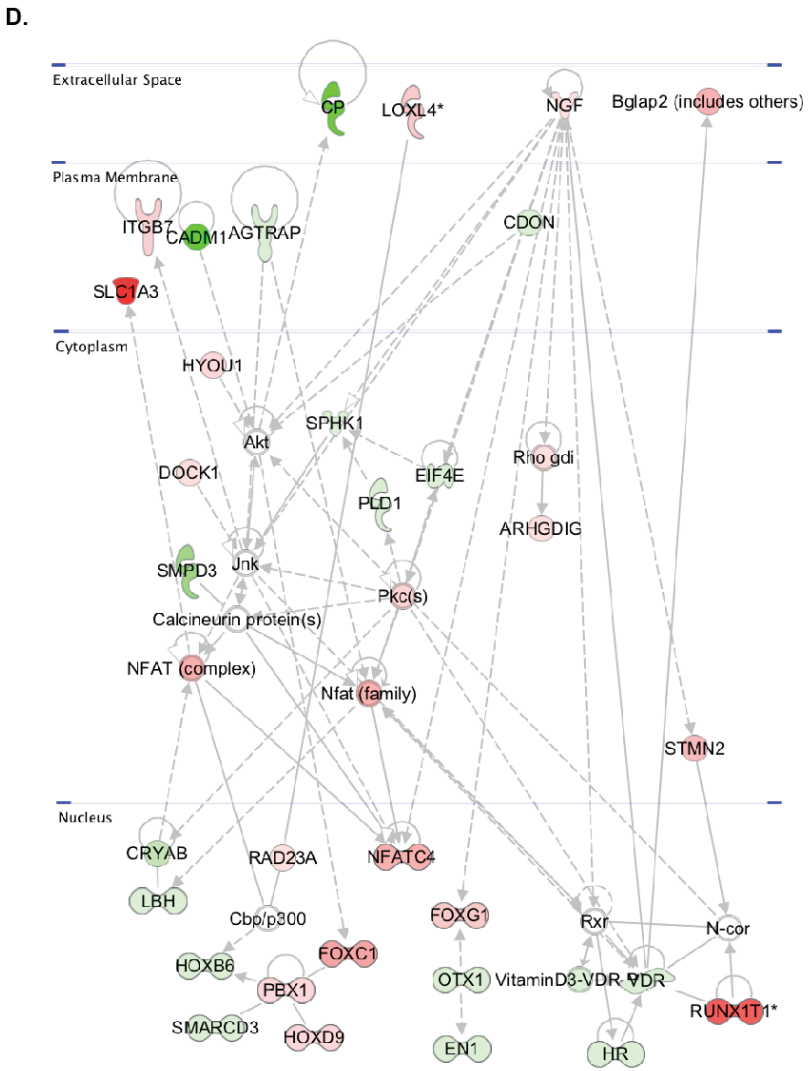
A.


Extracellular Space
HGF
IGFBP2
CLEC11A
Igfbp
IGFBP5
IGFBP4
SPP1
AREG/AREGB
IL11
VEGFA
IGF2
Plasma Membrane
HPSE
SCUBE3
Lh
H19
EZR
PODXL
FSH
Cytoplasm
RGS16
Vegf
hCG
ERK1/2
PRKD1
Nucleus
Alp
CCNG2
DUSP4
Creb
Hdac
MNT
ZNF440/ZNF808

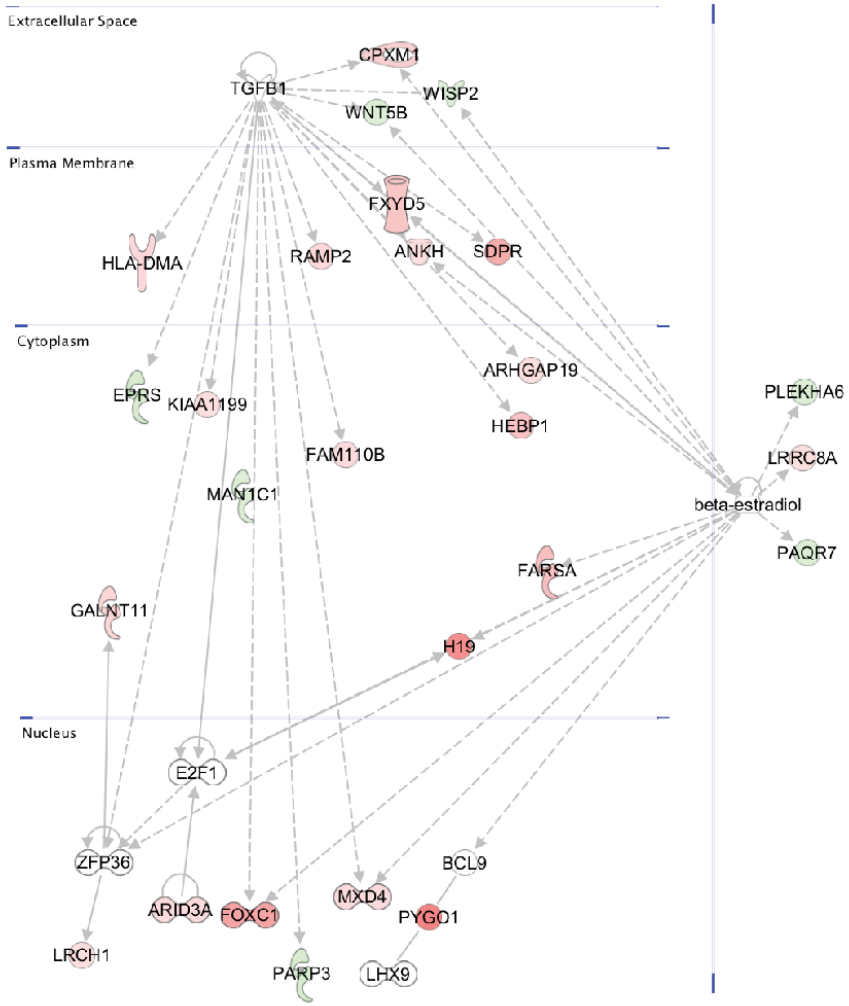
B.


Extracellular Space
PRELP
COL2A1
PPBP
COL15A1
WNT5B
WNT4
Fibrin
WNT6
Wnt
Plasminogen Activator
WNT10A
SERPINE2
PLAU
COL4A5
FXYD5
PRSS35
ITGA11
GREM1
MMP10
DIO3
COL8A1
TNC
DDR2
FGF5
PDGFB
collagen
Plasma Membrane
Pdgfr
growth factor receptor
THBD
DDR1
ANXA4
Cytoplasm
HMGCS1
ENPP1
PTGIS
TOM1
CBR3
Nucleus
GFER
TRIM69
KLF2
NFkB (complex)
RIPK4
DBP
DUSP2
NR1D1

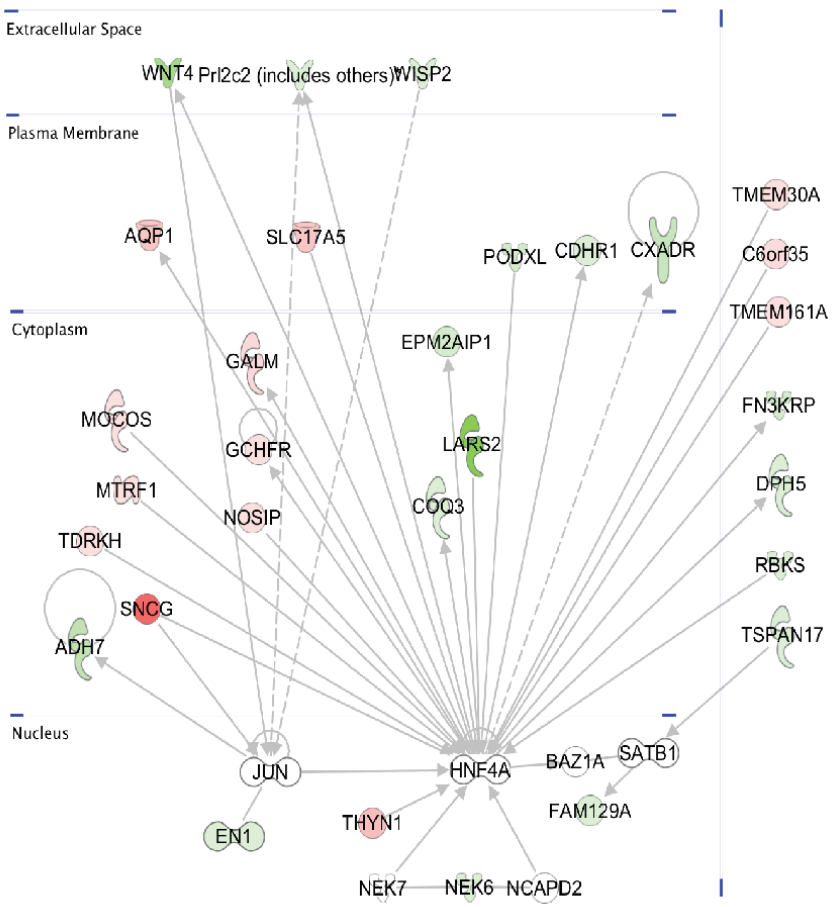

C.
OGN
IGFBP5
MMP2
AGT
IGFBP4
IL1B
CP
THSD1
DHRS7
AIM1
Plasma Membrane
GDPD3
ERBB2
THBD
Cytoplasm
PAIP1
CYP2E1
ARC
prostaglandin E2
H19
ST6GALNAC2
PTGIS
BCKDHB
RGS16
CHCHD10
PDE1A
DIO3
cyclic AMP
FAM134B
HEBP1
Rpl29 (includes others)
VGLL3
CHST2
CYP2S1
tretinoin
Cd59a
RCBTB2
Nucleus
IRX3
UAP1
CSRNP1


D.
Extracellular Space
CP
LOXL4*
NGF
Bglap2 (includes others)
Plasma Membrane
ITGB7
CADM1
AGTRAP
CDON
SLC1A3
Cytoplasm
HYOU1
SPHK1
Akt
DOCK1
EIF4E
Rho gdi
PLD1
ARHGDIG
SMPD3
Jnk
Pkc(s)
Calcineurin protein(s)
NFAT (complex)
Nfat (family)
STMN2
Nucleus
CRYAB
RAD23A
NFATC4
LBH
Cbp/p300
FOXG1
Rxr
N-cor
FOXC1
HOXB6
OTX1
VitaminD3-VDR
VDR
PBX1
RUNX1T1*
SMARCD3
HOXD9
EN1
HR

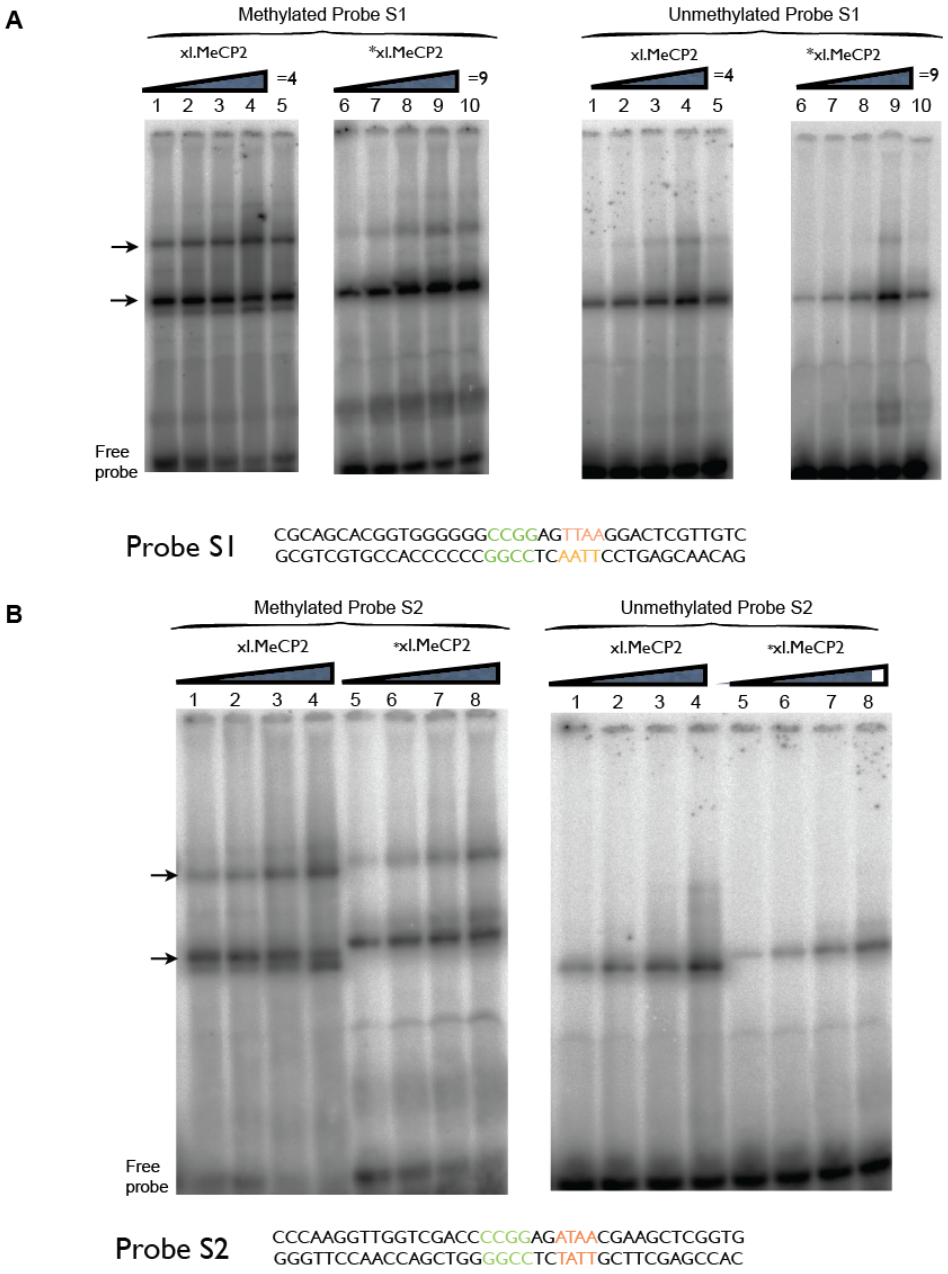
E.


Extracellular Space
Plasma Membrane
Cytoplasm
Nucleus
CPXM1
TGFB1
WISP2
WNT5B
FXYD5
HLA-DMA
RAMP2
ANKH
SDPR
ARHGAP19
EPRS
KIAA1199
HEBP1
PLEKHA6
FAM110B
LRRC8A
MAN1C1
beta-estradiol
PAQR7
FARSA
GALNT11
H19
E2F1
ZFP36
BCL9
ARID3A
FOXC1
MXD4
PYGO1
LRCH1
PARP3
LHX9

F.


Extracellular Space
Plasma Membrane
Cytoplasm
Nucleus
WNT4
Prl2c2 (includes others)
WISP2
AQP1
SLC17A5
PODXL
CDHR1
CXADR
TMEM30A
C6orf35
TMEM161A
EPM2AIP1
GALM
FN3KRP
MOCOS
GCHFR
LARS2
MTRF1
NOSIP
COQ3
DPH5
TDRKH
RBKS
SNCG
TSPAN17
ADH7
JUN
HNF4A
BAZ1A
SATB1
FAM129A
EN1
THYN1
NEK7
NEK6
NCAPD2

Supplementary Figure. 5.1. Significant protein clusters of misregulated transcripts from the virus transformed Mbd4$^{-/-}$ MEFs by IPA network analysis. (A) Cellular Movement, DNA Replication (Growth hormones related transcriptional regulation). (B) Cellular Movement, Cellular Growth and Proliferation (Collagen induced and NFkB associated). (C) Endocrine System Development and Function (Type II nuclear receptor RXR and retinoic acid associated). (D) Cellular Development, Cellular Assembly and Organization (Type II non-steroid nuclear receptor VDR related) (E) Cancer, Reproductive System Disease, Renal and Urological Disease (Steroid nuclear receptor ERα related). (F) Cellular Assembly and Organization, Cellular Function and Maintenance (orphan nuclear receptor HNF4a related). Note the proteins with red (upregulation) or green (downregulation) filling are the candidates from my mis-regulation gene list; the ones without color filling are those involved in the indicated pathway but not in my candidate list. The solid lines are direct relationships; and the dotted lines are indirect relationships. The arrows generally indicate the positive regulation. The annotations for symbols are in Supplementary Figure. 3.1.



Supplementary Figure. 6.1. The binding specificity of xl.MeCP2. Binding of the duplex probes containing 5-methylcytosine (5mC)(left) or the responsive unmethylated cytosines (right) in the context of CpG to varying concentrations of the methyl-binding domain (MBD) of (A. probe S1) and (B. probe S2) MeCP2 from 100 to 800 nM; sequence of duplex is seen in the Figure. The correct bands are indicated by black arrows. The procedure of radioactive EMSAs is in methods.

Supplementary Table 7.1 A

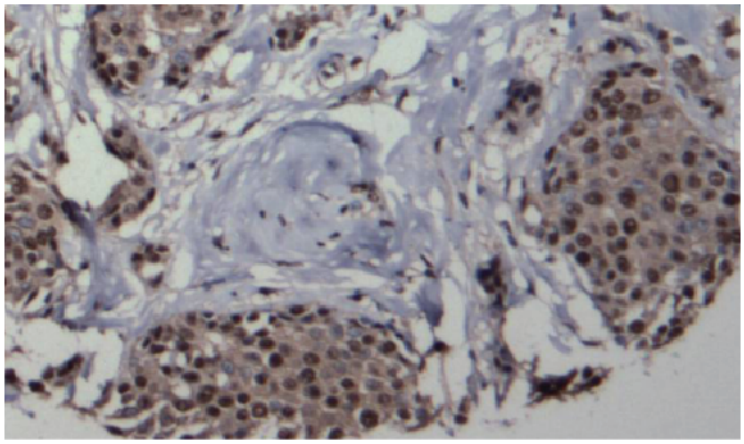
Gene	Promoter CGI	DNA meth HELP Seq MEF	DNA methy Meissner MEF	H3K27me3 in $p53^{-/-}$ MEF	H3K4me3 in $p53^{-/-}$ MEF	Expression in $p53^{-/-}$ MEF
Sox9	y	low	low	n	y	y
Sox2	y	low	low	y	y (low)	n
Sox11	y	low	low	n	y	y
Klf4	y	low	low	n	y	y
Klf2	y	low	low	n	y	y

Supplementary Table 7.1 B

SYMBOL	RNAPII	H3me3K9	H3me3K27	meC	ACCESSION	GENE_ID	Species
SOX9	0.8625	0.315	1.8125	0.3075	NM_000346.2	6662	human
KLF4	1.12	0.36	1.75	0.435	NM_004235.3	9314	human
KLF2	0.3175	0.215	0.825	0.6775	NM_016270.2	10365	human
CDKN1A	1.7925	0.485	1.2975	0.515	NM_078467.1	1026	human
SOX2	3.3125	0.4	0.9575	0.255	NM_003106.2	6657	human
SOX11	1.605	0.5325	0.5075	0.4825	NM_003108.3	6664	human
Sox9	1.5775	0.2425	0.045	0.34	NM_011448	20682	mouse
Klf4	1.66	0.15	0.0475	0.2275	NM_010637	16600	mouse
Klf2	0.6725	0.1075	0.5275	0.18	NM_008452	16598	mouse
Cdkn1a	1.7925	0.2375	0.1875	0.445	NM_007669	12575	mouse
Sox2	0.055	0.35	0.5625	0.9525	NM_011443	20674	mouse
Sox11	0.225	0.2525	0.6225	0.4225	Z18960	20666	mouse
Median	0.4825	0.36	0.425	0.415			human
	0.3975	0.2725	0.17	0.235			mouse

Supplementary Table 7.1. The methylation statuses of some developmental genes that are potential MBD4 targets are usually low (yellow labeled). A. DNA methylation at promoters of selected genes. The promoter CGIs were checked from UCSC browser. Other Data from (Meissner et al., 2008a) (DNA methy meissner MEF) and all the others are from unpublished data of James Reddington Thesis (Reddington, 2011). MEF: mouse embryonix fibrast; HELP: HpaII tiny fragment Enrichment by Ligation-mediated PCR. B. DNA methylation statuses of promoters of selected genes using (Komashko et al., 2008)'s data (Max4 enrichment values for all marks in all human and mouse cell lines). Briefly, Max4 is peak calling method (Komashko et al., 2008) that a value was assigned based on the highest mean of four consecutive probes in each promoter, then obtained results were combined if two-array designs were used (Komashko et al., 2008). Medians based on all the genes in Komashko's Max4 human and mouse tables were calculated, based on which low (yellow) or high (red) values were estimated. It worth noting that almost all the methylated cytosine value at the promoter regions of these genes in human and mouse cell lines are low, which consistent with table A.

A

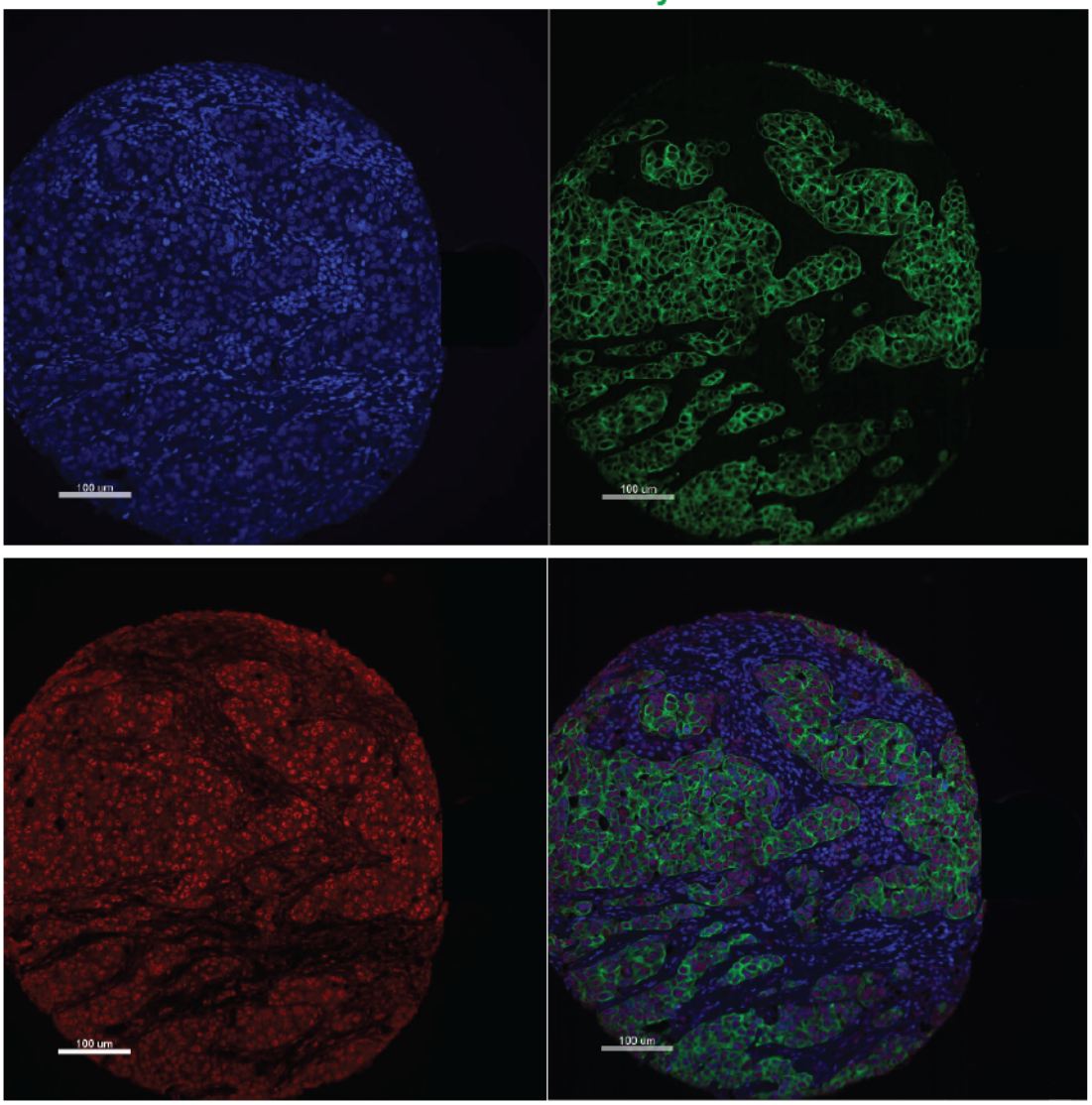


MBD4 IHC
brown staining

B

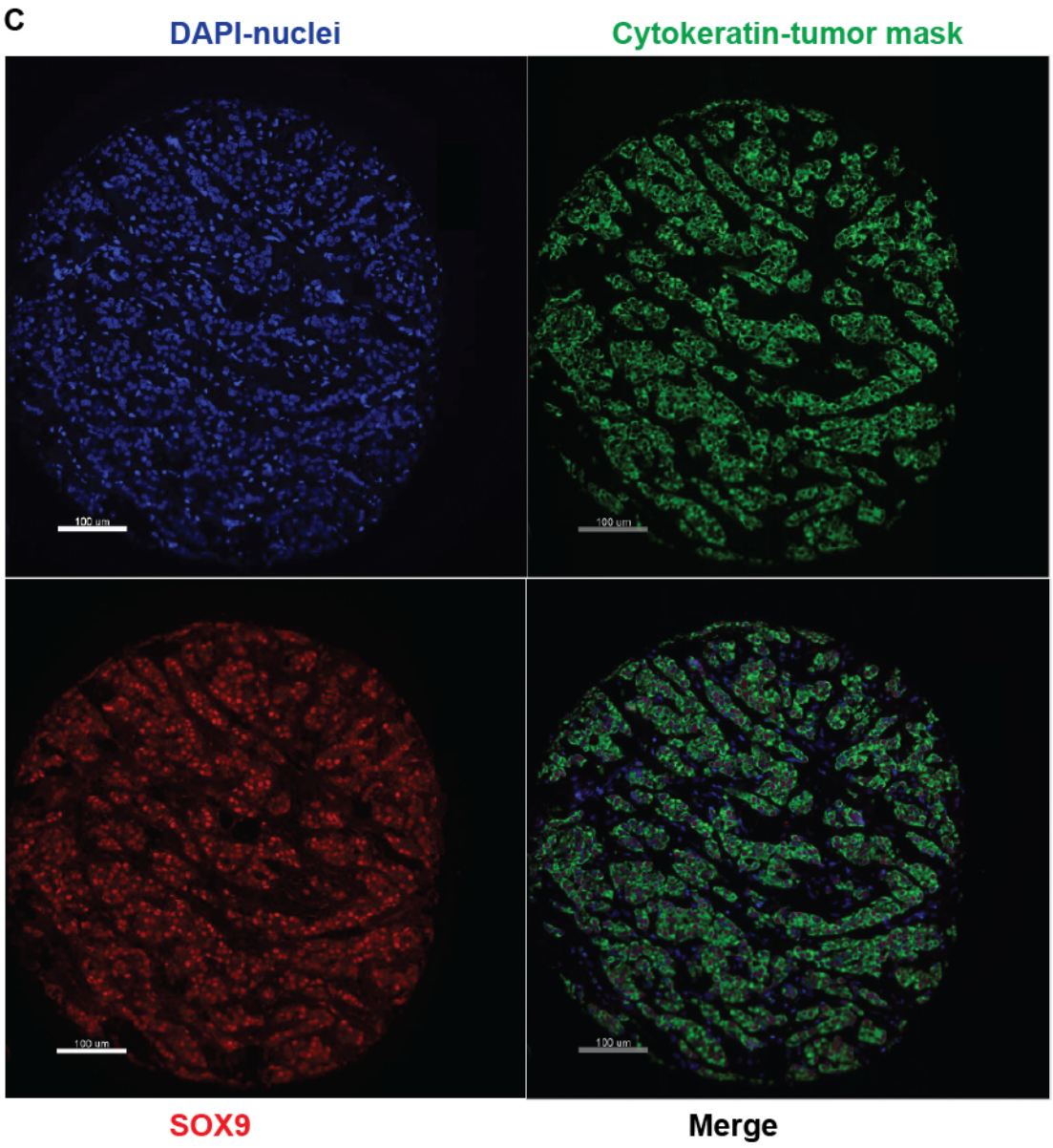
DAPI-nuclei

Cytokeratin-tumor mask



MBD4

Merge



Supplementary Figure. 7.1. MBD4 and SOX9 expression in primary breast tumor tissues. (A) Immunohistochemistry using brown staining for MBD4 in primary breast tumor tissues are shown. MBD4 is clearly expressing in both nuclear and cytoplasm (with approximate ratio at 2:1 for Nuclear:cytoplasm expression in most cases, unpublished data). The stroma area is clean. (B&C) AQUA quantitative image analysis for (B) MBD4 and (C) SOX9. DAPI counterstain was used to identify nuclei (B&C, top left), pan-cytokeratin antibody (epithelial marker) to identify infiltrating tumour cells (B&C, top right), and Cy-5–tyramide for detection of targets of (B) MBD4 and (C) SOX9 (B&C, bottom left) for compartmentalised analysis of tissue sections. The merge of all three stainings are shown in (B&C bottom right) Representative immunofluorescence images of tissue microarrays (TMA) for MBD4 and SOX9 are shown.

Bibliography

ABDEL-RAHMAN, W. M., KNUUTILA, S., PELTOMAKI, P., HARRISON, D. J. & BADER, S. A. 2008. Truncation of MBD4 predisposes to reciprocal chromosomal translocations and alters the response to therapeutic agents in colon cancer cells. *DNA Repair (Amst),* 7**,** 321-8.

ABRAMSON, J., GIRAUD, M., BENOIST, C. & MATHIS, D. 2010. Aire's partners in the molecular control of immunological tolerance. *Cell,* 140**,** 123-135.

AFONJA, O., RAAKA, B. M., HUANG, A., DAS, S., ZHAO, X., HELMER, E., JUSTE, D. & SAMUELS, H. H. 2002. RAR agonists stimulate SOX9 gene expression in breast cancer cell lines: evidence for a role in retinoid-mediated growth inhibition. *Oncogene,* 21**,** 7850-7860.

AIZAWA, K., SUZUKI, T., KADA, N., ISHIHARA, A., KAWAI-KOWASE, K., MATSUMURA, T., SASAKI, K., MUNEMASA, Y., MANABE, I., KURABAYASHI, M., COLLINS, T. & NAGAI, R. 2004. Regulation of platelet-derived growth factor-A chain by Krüppel-like factor 5: new pathway of cooperative activation with nuclear factor-kappaB. *The Journal of biological chemistry,* 279**,** 70-76.

ALBALAT, R. 2008. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus Branchiostoma floridae. *Dev Genes Evol,* 218**,** 691-701.

ALETAHA, D., NEOGI, T., SILMAN, A. J., FUNOVITS, J., FELSON, D. T., BINGHAM, C. O., BIRNBAUM, N. S., BURMESTER, G. R., BYKERK, V. P., COHEN, M. D., COMBE, B., COSTENBADER, K. H., DOUGADOS, M., EMERY, P., FERRACCIOLI, G., HAZES, J. M., HOBBS, K., HUIZINGA, T. W., KAVANAUGH, A., KAY, J., KVIEN, T. K., LAING, T., MEASE, P., MENARD, H. A., MORELAND, L. W., NADEN, R. L., PINCUS, T., SMOLEN, J. S., STANISLAWSKA-BIERNAT, E., SYMMONS, D., TAK, P. P., UPCHURCH, K. S., VENCOVSKY, J., WOLFE, F. & HAWKER, G. 2010. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Annals of the Rheumatic Diseases,* 69**,** 1580-1588.

ANCELIN, K., LANGE, U. C., HAJKOVA, P., SCHNEIDER, R., BANNISTER, A. J., KOUZARIDES, T. & SURANI, M. A. 2006. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol, 8, 623-30.ANDERSON, R. M., BOSCH, J. A., GOLL, M. G., HESSELSON, D., DONG, P. D., SHIN, D., CHI, N. C., SHIN, C. H., SCHLEGEL, A., HALPERN, M. & STAINIER, D. Y. 2009. Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Dev Biol,* 334**,** 213-23.

ANTEQUERA, F., BOYES, J. & BIRD, A. 1990. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell,* 62**,** 503-514.

ARAVIND, L., ABHIMAN, S. & IYER, L. M. 2011. *Chapter 3 - Natural History of the Eukaryotic Chromatin Protein Methylation System*, Elsevier Inc.

ARITA, K., ARIYOSHI, M., TOCHIO, H., NAKAMURA, Y. & SHIRAKAWA, M. 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature,* 455**,** 818-821.

ATKINS, G. B., WANG, Y., MAHABELESHWAR, G. H., SHI, H., GAO, H., KAWANAMI, D., NATESAN, V., LIN, Z., SIMON, D. I. & JAIN, M. K. 2008.

Hemizygous Deficiency of Kruppel-Like Factor 2 Augments Experimental Atherosclerosis. *Circulation Research,* 103**,** 690-693.

AVVAKUMOV, G. V., WALKER, J. R., XUE, S., LI, Y., DUAN, S., BRONNER, C., ARROWSMITH, C. H. & DHE-PAGANON, S. 2008. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature,* 455**,** 822-825.

AYYANATHAN, K. 2003. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes & Development,* 17**,** 1855-1869.

BADER, S., WALKER, M., HENDRICH, B., BIRD, A., BIRD, C., HOOPER, M. & WYLLIE, A. 1999. Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene,* 18**,** 8044-7.

BADER, S. A., WALKER, M. & HARRISON, D. J. 2007. A human cancer-associated truncation of MBD4 causes dominant negative impairment of DNA repair in colon cancer cells. *Br J Cancer,* 96**,** 660-6.

BAKER, R. T., CATANZARITI, A.-M., KARUNASEKARA, Y., SOBOLEVA, T. A., SHARWOOD, R., WHITNEY, S. & BOARD, P. G. 2005. Using deubiquitylating enzymes as research tools. *Methods in enzymology,* 398**,** 540-554.

BALLESTAR, E., PAZ, M. F., VALLE, L., WEI, S., FRAGA, M. F., ESPADA, J., CIGUDOSA, J. C., HUANG, T. H.-M. & ESTELLER, M. 2003. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *The EMBO Journal,* 22**,** 6335-6345.

BALLESTAR, E. & WOLFFE, A. P. 2001. Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur J Biochem,* 268**,** 1-6.

BARANZINI, S. E., MUDGE, J., VAN VELKINBURGH, J. C., KHANKHANIAN, P., KHREBTUKOVA, I., MILLER, N. A., ZHANG, L., FARMER, A. D., BELL, C. J., KIM, R. W., MAY, G. D., WOODWARD, J. E., CAILLIER, S. J., MCELROY, J. P., GOMEZ, R., PANDO, M. J., CLENDENEN, L. E., GANUSOVA, E. E., SCHILKEY, F. D., RAMARAJ, T., KHAN, O. A., HUNTLEY, J. J., LUO, S., KWOK, P. Y., WU, T. D., SCHROTH, G. P., OKSENBERG, J. R., HAUSER, S. L. & KINGSMORE, S. F. 2010. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature,* 464**,** 1351-6.

BARDE, I., LAURENTI, E., VERP, S., GRONER, A. C., TOWNE, C., PADRUN, V., AEBISCHER, P., TRUMPP, A. & TRONO, D. 2009. Regulation of Episomal Gene Expression by KRAB/KAP1-Mediated Histone Modifications. *Journal of Virology,* 83**,** 5574-5580.

BARRERO, M. J. & MALIK, S. 2006. Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. Mol Cell, 24, 233-43.BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

BARTEL, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* 116**,** 281-97.

BARTKE, T., VERMEULEN, M., XHEMALCE, B., ROBSON, S. C., MANN, M. & KOUZARIDES, T. 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell,* 143**,** 470-84.

BEAUJEAN, N., TAYLOR, J., GARDNER, J., WILMUT, I., MEEHAN, R. & YOUNG, L. 2004. Effect of limited DNA methylation reprogramming in the normal sheep embryo on somatic cell nuclear transfer. *Biology of reproduction,* 71**,** 185-93.

BECKER, P. B. & HORZ, W. 2002. ATP-dependent nucleosome remodeling. *Annual review of biochemistry,* 71**,** 247-73.

BEER, D. G., KARDIA, S. L. R., HUANG, C.-C., GIORDANO, T. J., LEVIN, A. M., MISEK, D. E., LIN, L., CHEN, G., GHARIB, T. G., THOMAS, D. G., LIZYNESS,

M. L., KUICK, R., HAYASAKA, S., TAYLOR, J. M. G., IANNETTONI, M. D., ORRINGER, M. B. & HANASH, S. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*.

BELLACOSA, A., CICCHILLITTI, L., SCHEPIS, F., RICCIO, A., YEUNG, A. T., MATSUMOTO, Y., GOLEMIS, E. A., GENUARDI, M. & NERI, G. 1999. MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 3969-74.

BESTOR, T. H. 2000. The DNA methyltransferases of mammals. *Human molecular genetics,* 9**,** 2395-402.

BICKLE, T. A. & KRUGER, D. H. 1993. Biology of DNA restriction. *Microbiol Rev,* 57**,** 434-50.

BOLAND, M. J. & CHRISTMAN, J. K. 2008. Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *Journal of molecular biology,* 379**,** 492-504.

BOSTICK, M., KIM, J. K., ESTEVE, P. O., CLARK, A., PRADHAN, S. & JACOBSEN, S. E. 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science (New York, NY),* 317**,** 1760-1764.

BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature,* 441**,** 349-53.

BRAND, D. D., LATHAM, K. A. & ROSLONIEC, E. F. 2007. Collagen-induced arthritis. *Nature protocols,* 2**,** 1269-1275.

BRIERS, S., CRAWFORD, C., BICKMORE, W. A. & SUTHERLAND, H. G. 2009. KRAB zinc-finger proteins localise to novel KAP1-containing foci that are adjacent to PML nuclear bodies. *Journal of Cell Science,* 122**,** 937-946.

BRÜCKNER, A., POLGE, C., LENTZE, N., AUERBACH, D. & SCHLATTNER, U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *International journal of molecular sciences,* 10**,** 2763-2788.

BURDZY, A., NOYES, K. T., VALINLUCK, V. & SOWERS, L. C. 2002. Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA. *Nucleic acids research,* 30**,** 4068-74.

CAMMAS, F., MARK, M., DOLLÉ, P., DIERICH, A., CHAMBON, P. & LOSSON, R. 2000. Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development,* 127**,** 2955-2963.

CAMPANERO, M. R., ARMSTRONG, M. I. & FLEMINGTON, E. K. 2000. CpG methylation as a mechanism for the regulation of E2F activity. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 6481-6.

CEDAR, H. & BERGMAN, Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews. Genetics,* 10**,** 295-304.

CESARO, E., DE CEGLI, R., MEDUGNO, L., FLORIO, F., GROSSO, M., LUPO, A., IZZO, P. & COSTANZO, P. 2009. The Kruppel-like zinc finger protein ZNF224 recruits the arginine methyltransferase PRMT5 on the transcriptional repressor complex of the aldolase A gene. *Journal of Biological Chemistry,* 284**,** 32321-32330.

CHAWLA, A. 2001. Nuclear Receptors and Lipid Physiology: Opening the X-Files. *Science (New York, NY),* 294**,** 1866-1870.

CHEN, D., LUCEY, M. J., PHOENIX, F., LOPEZ-GARCIA, J., HART, S. M., LOSSON, R., BULUWELA, L., COOMBES, R. C., CHAMBON, P., SCHÄR, P. & ALI, S. 2003. T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of

estrogen-regulated genes through direct interaction with estrogen receptor alpha. *The Journal of biological chemistry,* 278**,** 38586-38592.

CHEN, T., HEVI, S., GAY, F., TSUJIMOTO, N., HE, T., ZHANG, B., UEDA, Y. & LI, E. 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature genetics,* 39**,** 391-6.

CHEUNG, M., CHABOISSIER, M.-C., MYNETT, A., HIRST, E., SCHEDL, A. & BRISCOE, J. 2005. The Transcriptional Control of Trunk Neural Crest Induction, Survival, and Delamination. *Developmental Cell,* 8**,** 179-192.

CHINNUSAMY, V. & ZHU, J. K. 2009. Rna-directed DNA methylation and demethylation in plants. Science in China, Series C: Life Sciences, 52, 331-343.

CLAPIER, C. R. & CAIRNS, B. R. 2009. The biology of chromatin remodeling complexes. *Annual Review of Biochemistry,* 78**,** 273-304.

CLOUAIRE, T., DE LAS HERAS, J. I., MERUSI, C. & STANCHEVA, I. 2010. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic acids research,* 38**,** 4620-34.

COLLINS, N., POOT, R. A., KUKIMOTO, I., GARCÍA-JIMÉNEZ, C., DELLAIRE, G. & VARGA-WEISZ, P. D. 2002. An ACF1–ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nature Genetics,* 32**,** 627-632.

CORTAZAR, D., KUNZ, C., SELFRIDGE, J., LETTIERI, T., SAITO, Y., MACDOUGALL, E., WIRZ, A., SCHUERMANN, D., JACOBS, A. L., SIEGRIST, F., STEINACHER, R., JIRICNY, J., BIRD, A. & SCHAR, P. 2011. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature,* 470**,** 419-23.

CORTELLINO, S., TURNER, D., MASCIULLO, V., SCHEPIS, F., ALBINO, D., DANIEL, R., SKALKA, A. M., MEROPOL, N. J., ALBERTI, C., LARUE, L. & BELLACOSA, A. 2003. The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 15071-6.

CORTELLINO, S., XU, J., SANNAI, M., MOORE, R., CARETTI, E., CIGLIANO, A., LE COZ, M., DEVARAJAN, K., WESSELS, A., SOPRANO, D., ABRAMOWITZ, L. K., BARTOLOMEI, M. S., RAMBOW, F., BASSI, M. R., BRUNO, T., FANCIULLI, M., RENNER, C., KLEIN-SZANTO, A. J., MATSUMOTO, Y., KOBI, D., DAVIDSON, I., ALBERTI, C., LARUE, L. & BELLACOSA, A. 2011. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell,* 146**,** 67-79.

COUREY, A. J. & JIA, S. 2001. Transcriptional repression: the long and the short of it. *Genes & Development,* 15**,** 2786-2796.

CROSS, N. C. P. 2011. Genetic and epigenetic complexity in myeloproliferative neoplasms. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program,* 2011**,** 208-214.

CUMMINS, J. M., RAGO, C., KOHLI, M., KINZLER, K. W., LENGAUER, C. & VOGELSTEIN, B. 2004. Tumour suppression: Disruption of HAUSP gene stabilizes p53. *Nature Cell Biology,* 428.

CUMMINS, J. M. & VOGELSTEIN, B. 2004. HAUSP is required for p53 destabilization. *Cell cycle (Georgetown, Tex),* 3**,** 689-692.

DAS, S., FOLEY, N., BRYAN, K., WATTERS, K. M., BRAY, I., MURPHY, D. M., BUCKLEY, P. G. & STALLINGS, R. L. 2010. MicroRNA mediates DNA demethylation events triggered by retinoic acid during neuroblastoma cell differentiation. *Cancer research,* 70**,** 7874-81.

DEANGELIS, J. T. 2010. AN INVESTIGATION OF THE MOLECULAR MECHANISMS INVOLVED IN THE GENERATION OF BREAST CANCER. *PhD thesis, The University of Alabama at Birmingham*.

DHAWAN, S., GEORGIA, S., TSCHEN, S. I., FAN, G. & BHUSHAN, A. 2011. Pancreatic beta cell identity is maintained by DNA methylation-mediated repression of Arx. *Dev Cell,* 20**,** 419-29.

DHAYALAN, A., RAJAVELU, A., RATHERT, P., TAMAS, R., JURKOWSKA, R. Z., RAGOZIN, S. & JELTSCH, A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *The Journal of biological chemistry,* 285**,** 26114-20.

DIALYNAS, G. K., VITALINI, M. W. & WALLRATH, L. L. 2008. Linking Heterochromatin Protein 1 (HP1) to cancer progression. *Mutation research,* 647**,** 13-20.

DOERFLER, W. 1983. DNA methylation and gene activity. *Annual Review of Biochemistry,* 52**,** 93-124.

DOERFLER, W. 2006. De novo methylation, long-term promoter silencing, methylation patterns in the human genome, and consequences of foreign DNA insertion. *Curr Top Microbiol Immunol,* 301**,** 125-75.

DRIVDAHL, R., HAUGK, K. H., SPRENGER, C. C., NELSON, P. S., TENNANT, M. K. & PLYMATE, S. R. 2004. Suppression of growth and tumorigenicity in the prostate tumor cell line M12 by overexpression of the transcription factor SOX9. *Oncogene,* 23**,** 4584-4593.

DU, Z., SONG, J., WANG, Y., ZHAO, Y., GUDA, K., YANG, S., KAO, H.-Y., XU, Y., WILLIS, J., MARKOWITZ, S. D., SEDWICK, D., EWING, R. M. & WANG, Z. 2010. DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Science Signaling,* 3**,** ra80.

DUNCAN, B. K. & MILLER, J. H. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature,* 287**,** 560-1.

DUNICAN, D. S., RUZOV, A., HACKETT, J. A. & MEEHAN, R. R. 2008. xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function. *Development,* 135**,** 1295-302.

EBERHARTER, A., VETTER, I., FERREIRA, R. & BECKER, P. B. 2004. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *The EMBO Journal,* 23**,** 4029-4039.

EHRLICH, M. & WANG, R. Y. 1981. 5-Methylcytosine in eukaryotic DNA. *Science (New York, NY),* 212**,** 1350-1357.

ESTEVE, P. O., CHANG, Y., SAMARANAYAKE, M., UPADHYAY, A. K., HORTON, J. R., FEEHERY, G. R., CHENG, X. & PRADHAN, S. 2011. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nature structural & molecular biology,* 18**,** 42-8.

ESTEVE, P. O., CHIN, H. G., BENNER, J., FEEHERY, G. R., SAMARANAYAKE, M., HORWITZ, G. A., JACOBSEN, S. E. & PRADHAN, S. 2009. Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 5076-81.

ESTEVE, P. O., CHIN, H. G., SMALLWOOD, A., FEEHERY, G. R., GANGISETTY, O., KARPF, A. R., CAREY, M. F. & PRADHAN, S. 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & development,* 20**,** 3089-103.

EWING, A. K., ATTNER, M. & CHAKRAVARTI, D. 2007. Novel Regulatory Role for Human Acf1 in Transcriptional Repression of Vitamin D3 Receptor-Regulated Genes. *Molecular endocrinology (Baltimore, Md.),* 21**,** 1791-1806.

FELDMAN, N., GERSON, A., FANG, J., LI, E., ZHANG, Y., SHINKAI, Y., CEDAR, H. & BERGMAN, Y. 2006. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature cell biology,* 8**,** 188-94.

FELLE, M., JOPPIEN, S., NÉMETH, A., DIERMEIER, S., THALHAMMER, V., DOBNER, T., KREMMER, E., KAPPLER, R. & LÄNGST, G. 2011. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Research*.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, NY),* 246**,** 64-71.

FERGUSON-SMITH, A. C. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews Genetics,* 12**,** 565-575.

FIELDS, S. & SONG, O. 1989. A novel genetic system to detect protein-protein interactions. *Nature,* 340**,** 245-246.

FIGUEROA, M. E., ABDEL-WAHAB, O., LU, C., WARD, P. S., PATEL, J., SHIH, A., LI, Y., BHAGWAT, N., VASANTHAKUMAR, A., FERNANDEZ, H. F., TALLMAN, M. S., SUN, Z., WOLNIAK, K., PEETERS, J. K., LIU, W., CHOE, S. E., FANTIN, V. R., PAIETTA, E., LOWENBERG, B., LICHT, J. D., GODLEY, L. A., DELWEL, R., VALK, P. J., THOMPSON, C. B., LEVINE, R. L. & MELNICK, A. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer cell,* 18**,** 553-67.

FILION, G. J., ZHENILO, S., SALOZHIN, S., YAMADA, D., PROKHORTCHOUK, E. & DEFOSSEZ, P. A. 2006. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and cellular biology,* 26**,** 169-81.

FIONA MACDOUGALL, E. 2007. Functional analysis of the DNA repair protein MBD4.

FIRESTEIN, G. S. 2003. Evolving concepts of rheumatoid arthritis. *Nature,* 423**,** 356-361.

FLORIO, F., CESARO, E., MONTANO, G., IZZO, P., MILES, C. & COSTANZO, P. 2010. Biochemical and functional interaction between ZNF224 and ZNF255, two members of the Kruppel-like zinc-finger protein family and WT1 protein isoforms. *Human molecular genetics,* 19**,** 3544-56.

FOSTER, K. W., FROST, A. R., MCKIE-BELL, P., LIN, C. Y., ENGLER, J. A., GRIZZLE, W. E. & RUPPERT, J. M. 2000. Increase of GKLF messenger RNA and protein expression during progression of breast cancer. *Cancer Research,* 60**,** 6488-6495.

FOUSE, S. D., SHEN, Y., PELLEGRINI, M., COLE, S., MEISSNER, A., VAN NESTE, L., JAENISCH, R. & FAN, G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell stem cell,* 2**,** 160-9.

FRAGA, M. F., BALLESTAR, E., MONTOYA, G., TAYSAVANG, P., WADE, P. A. & ESTELLER, M. 2003. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Research,* 31**,** 1765-1774.

FRAUER, C., HOFFMANN, T., BULTMANN, S., CASA, V., CARDOSO, M. C., ANTES, I. & LEONHARDT, H. 2011. Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. *PLoS ONE,* 6**,** e21306.

FREE, A., WAKEFIELD, R. I., SMITH, B. O., DRYDEN, D. T., BARLOW, P. N. & BIRD, A. P. 2001. DNA recognition by the methyl-CpG binding domain of MeCP2. *The Journal of biological chemistry,* 276**,** 3353-3360.

FUKS, F., HURD, P. J., WOLF, D., NAN, X., BIRD, A. P. & KOUZARIDES, T. 2003. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *The Journal of biological chemistry,* 278**,** 4035-40.

FUKUSHIGE, S., KONDO, E., GU, Z., SUZUKI, H. & HORII, A. 2006. RET finger protein enhances MBD2- and MBD4-dependent transcriptional repression. *Biochemical and Biophysical Research Communications,* 351**,** 85-92.

GASTON, K. & JAYARAMAN, P. S. 2003. Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cellular and molecular life sciences : CMLS,* 60**,** 721-741.

GLOECKNER, C. J., BOLDT, K., SCHUMACHER, A., ROEPMAN, R. & UEFFING, M. 2007. A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *PROTEOMICS,* 7**,** 4228-4234.

GOLL, M. G. & BESTOR, T. H. 2005. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry,* 74**,** 481-514.

GORDON, C. T., TAN, T. Y., BENKO, S., FITZPATRICK, D., LYONNET, S. & FARLIE, P. G. 2009. Long-range regulation at the SOX9 locus in development and disease. *Journal of medical genetics,* 46**,** 649-656.

GRONER, A. C., MEYLAN, S., CIUFFI, A., ZANGGER, N., AMBROSINI, G., DÉNERVAUD, N., BUCHER, P. & TRONO, D. 2010. KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. *PLoS Genetics,* 6**,** e1000869.

GUO, G., WANG, W. & BRADLEY, A. 2004. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature,* 429**,** 891-5.

GUO, W., KECKESOVA, Z., DONAHER, J. L., SHIBUE, T., TISCHLER, V., REINHARDT, F., ITZKOVITZ, S., NOSKE, A., ZÜRRER-HÄRDI, U., BELL, G., TAM, W. L., MANI, S. A., VAN OUDENAARDEN, A. & WEINBERG, R. A. 2012. Slug and Sox9 Cooperatively Determine the Mammary Stem Cell State. *Cell,* 148**,** 1015-1028.

Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science 286(5441): 950-952.

HARIKRISHNAN, K. N., CHOW, M. Z., BAKER, E. K., PAL, S., BASSAL, S., BRASACCHIO, D., WANG, L., CRAIG, J. M., JONES, P. L., SIF, S. & EL-OSTA, A. 2005. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nature genetics,* 37**,** 254-64.

HASHIMOTO, H., HORTON, J. R., ZHANG, X., BOSTICK, M., JACOBSEN, S. E. & CHENG, X. 2008. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature,* 455**,** 826-829.

HASHIMOTO, H., ZHANG, X. & CHENG, X. 2012. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. Nucleic Acids Res, 40, 8276-84.

HENDRICH, B. & BIRD, A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol,* 18**,** 6538-47.

HENDRICH, B., HARDELAND, U., NG, H. H., JIRICNY, J. & BIRD, A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature,* 401**,** 301-4.

HENDRICH, B. & TWEEDIE, S. 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet,* 19**,** 269-77.

HERVOUET, E., VALLETTE, F. M. & CARTRON, P. F. 2009. Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics : official journal of the DNA Methylation Society,* 4**,** 487-99.

HODGES, E., SMITH, A. D., KENDALL, J., XUAN, Z., RAVI, K., ROOKS, M., ZHANG, M. Q., YE, K., BHATTACHARJEE, A., BRIZUELA, L., MCCOMBIE, W. R., WIGLER, M., HANNON, G. J. & HICKS, J. B. 2009. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome research,* 19**,** 1593-605.

HOLLIDAY, R. & PUGH, J. E. 1975. DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-32.

HOLOWATY, M. N. 2003. Protein Interaction Domains of the Ubiquitin-specific Protease, USP7/HAUSP. *Journal of Biological Chemistry,* 278**,** 47753-47761.

HOTCHKISS, R. D. 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry,* 175**,** 315-32.

HOWARD, J. H., FROLOV, A., TZENG, C.-W. D., STEWART, A., MIDZAK, A., MAJMUNDAR, A., GODWIN, A., HESLIN, M., BELLACOSA, A. & ARNOLETTI, J. P. 2009. Epigenetic downregulation of the DNA repair gene MED1/MBD4 in colorectal and ovarian cancer. *Cancer biology & therapy,* 8**,** 94-100.

HU, K., NAN, X., BIRD, A. & WANG, W. 2006. Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex. *Nature genetics,* 38**,** 962-4; author reply 964-7.

HUANG, C.-M., HUANG, P.-H., CHEN, C.-L., WAN, L., TSAI, C.-H., LIU, S.-C., HUANG, W.-L. & TSAI, F.-J. 2010a. MBD4 gene is associated with rheumatoid arthritis in Chinese patients in Taiwan. *Rheumatology international*.

HUANG, Y., PASTOR, W. A., SHEN, Y., TAHILIANI, M., LIU, D. R. & RAO, A. 2010b. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS one,* 5**,** e8888.

IGUCHI-ARIGA, S. M. & SCHAFFNER, W. 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & development,* 3**,** 612-9.

ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. *FEBS letters,* 583**,** 1713-20.

ILLINGWORTH, R. S., GRUENEWALD-SCHNEIDER, U., WEBB, S., KERR, A. R., JAMES, K. D., TURNER, D. J., SMITH, C., HARRISON, D. J., ANDREWS, R. & BIRD, A. P. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS genetics,* 6.

ITO, T., LEVENSTEIN, M. E., FYODOROV, D. V., KUTACH, A. K., KOBAYASHI, R. & KADONAGA, J. T. 1999. ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes & Development,* 13**,** 1529-1539.

IYENGAR, S. & FARNHAM, P. J. 2011. KAP1 Protein: An Enigmatic Master Regulator of the Genome. *Journal of Biological Chemistry,* 286**,** 26267-26276.

IYER, L. M., TAHILIANI, M., RAO, A. & ARAVIND, L. 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle,* 8**,** 1698-710.

JACKSON-GRUSBY, L., BEARD, C., POSSEMATO, R., TUDOR, M., FAMBROUGH, D., CSANKOVSZKI, G., DAUSMAN, J., LEE, P., WILSON, C., LANDER, E. & JAENISCH, R. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature genetics,* 27**,** 31-9.

JACOBS, A. L. & SCHÄR, P. 2011. DNA glycosylases: in DNA repair and beyond. *Chromosoma,* 121**,** 1-20.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics,* 33 Suppl**,** 245-54.

JAKOBSSON, J., CORDERO, M. I., BISAZ, R., GRONER, A. C., BUSSKAMP, V., BENSADOUN, J.-C., CAMMAS, F., LOSSON, R., MANSUY, I. M., SANDI, C. & TRONO, D. 2008. KAP1-Mediated Epigenetic Repression in the Forebrain Modulates Behavioral Vulnerability to Stress. *Neuron,* 60**,** 818-831.

JIANG, S. S., FANG, W. T., HOU, Y. H., HUANG, S. F., YEN, B. L., CHANG, J. L., LI, S. M., LIU, H. P., LIU, Y. L., HUANG, C. T., LI, Y. W., JANG, T. H., CHAN, S. H., YANG, S. J., HSIUNG, C. A., WU, C. W., WANG, L. H. & CHANG, I. S. 2010. Upregulation of SOX9 in Lung Adenocarcinoma and Its Involvement in the Regulation of Cell Growth and Tumorigenicity. *Clinical Cancer Research,* 16**,** 4363-4373.

JIN, S. G., KADAM, S. & PFEIFER, G. P. 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic acids research,* 38**,** e125.

JOHNSON, T. B. & COGHILL, R. D. 1925. RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS1. *Journal of the American Chemical Society,* 47**,** 2838-2844.

JONES, P. L., VEENSTRA, G. J., WADE, P. A., VERMAAK, D., KASS, S. U., LANDSBERGER, N., STROUBOULIS, J. & WOLFFE, A. P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics,* 19**,** 187-91.

JURKOWSKA, R. Z., JURKOWSKI, T. P. & JELTSCH, A. 2010. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem,* 12**,** 206-222.

KADAM, S. & EMERSON, B. M. 2003. Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Molecular Cell,* 11**,** 377-389.

KALUDOV, N. K. & WOLFFE, A. P. 2000. MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Research,* 28**,** 1921-1928.

KANEDA, M., HIRASAWA, R., CHIBA, H., OKANO, M., LI, E. & SASAKI, H. 2010. Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes to cells : devoted to molecular & cellular mechanisms*.

KANGASPESKA, S., STRIDE, B., METIVIER, R., POLYCARPOU-SCHWARZ, M., IBBERSON, D., CARMOUCHE, R. P., BENES, V., GANNON, F. & REID, G. 2008. Transient cyclical methylation of promoter DNA. *Nature,* 452**,** 112-5.

KATO, Y., KANEDA, M., HATA, K., KUMAKI, K., HISANO, M., KOHARA, Y., OKANO, M., LI, E., NOZAKI, M. & SASAKI, H. 2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human molecular genetics,* 16**,** 2272-80.

KAWASAKI, H. & TAIRA, K. 2004. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature,* 431**,** 211-7.

KHORASANIZADEH, S. 2004. The nucleosome: from genomic organization to genomic regulation. *Cell,* 116**,** 259-72.

KHORONENKOVA, S. V., DIANOVA, I. I., PARSONS, J. L. & DIANOV, G. L. 2011. USP7/HAUSP stimulates repair of oxidative DNA lesions. *Nucleic Acids Research,* 39**,** 2604-2609.

KIM, J. K., SAMARANAYAKE, M. & PRADHAN, S. 2009a. Epigenetic mechanisms in mammals. *Cellular and molecular life sciences : CMLS,* 66**,** 596-612.

KIM, M.-S., KONDO, T., TAKADA, I., YOUN, M.-Y., YAMAMOTO, Y., TAKAHASHI, S., MATSUMOTO, T., FUJIYAMA, S., SHIRODE, Y., YAMAOKA, I., KITAGAWA, H., TAKEYAMA, K.-I., SHIBUYA, H., OHTAKE, F. & KATO, S. 2009b. DNA demethylation in hormone-induced transcriptional derepression. *Nature,* 461**,** 1007-1012.

KIMURA, H. 2002. Methyl-CpG-binding Protein, MeCP2, Is a Target Molecule for Maintenance DNA Methyltransferase, Dnmt1. *Journal of Biological Chemistry,* 278**,** 4806-4812.

KIZILDAG, S., ATES, H. & KIZILDAG, S. 2010. Treatment of K562 cells with 1,25-dihydroxyvitamin D3 induces distinct alterations in the expression of apoptosis-related genes BCL2, BAX, BCLXL, and p21. Ann Hematol, 89, 1-7.KLOSE, R. J. & BIRD, A. P. 2004. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *The Journal of biological chemistry,* 279**,** 46490-6.

KLOSE, R. J., SARRAF, S. A., SCHMIEDEBERG, L., MCDERMOTT, S. M., STANCHEVA, I. & BIRD, A. P. 2005. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Molecular cell,* 19**,** 667-78.

KOKURA, K., KAUL, S. C., WADHWA, R., NOMURA, T., KHAN, M. M., SHINAGAWA, T., YASUKAWA, T., COLMENARES, C. & ISHII, S. 2001. The Ski protein family is required for MeCP2-mediated transcriptional repression. *The Journal of biological chemistry,* 276**,** 34115-21.

KOMASHKO, V. M., ACEVEDO, L. G., SQUAZZO, S. L., IYENGAR, S. S., RABINOVICH, A., O'GEEN, H., GREEN, R. & FARNHAM, P. J. 2008. Using ChIP-chip technology to reveal common principles of transcriptional repression in normal and cancer cells. *Genome Res,* 18**,** 521-32.

KON, N., KOBAYASHI, Y., LI, M., BROOKS, C. L., LUDWIG, T. & GU, W. 2010. Inactivation of HAUSP in vivo modulates p53 function. *Oncogene,* 29**,** 1270-9.

KONDO, E., GU, Z., HORII, A. & FUKUSHIGE, S. 2005. The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Molecular and Cellular Biology,* 25**,** 4388-4396.

KRAUSE, C. D., YANG, Z.-H., KIM, Y.-S., LEE, J.-H., COOK, J. R. & PESTKA, S. 2007. Protein arginine methyltransferases: Evolution and assessment of their pharmacological and therapeutic potential. *Pharmacology & therapeutics,* 113**,** 50-87.

KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-30.

KUNES, P., HOLUBCOVÁ, Z. & KREJSEK, J. 2009. Occurrence and significance of the nuclear transcription factor Krüppel-like factor 4 (KLF4) in the vessel wall. *Acta medica (Hradec Králové) / Universitas Carolina, Facultas Medica Hradec Králové,* 52**,** 135-139.

LAGET, S., JOULIE, M., LE MASSON, F., SASAI, N., CHRISTIANS, E., PRADHAN, S., ROBERTS, R. J. & DEFOSSEZ, P. A. 2010. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PloS one,* 5**,** e11982.

LAN, L., UI, A., NAKAJIMA, S., HATAKEYAMA, K., HOSHI, M., WATANABE, R., JANICKI, S. M., OGIWARA, H., KOHNO, T., KANNO, S.-I. & YASUI, A. 2010. The ACF1 Complex Is Required for DNA Double-Strand Break Repair in Human Cells. *Molecular Cell,* 40**,** 976-987.

LANGST, G. & BECKER, P. B. 2001. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *Journal of Cell Science,* 114**,** 2561-2568.

LAURENT, L., WONG, E., LI, G., HUYNH, T., TSIRIGOS, A., ONG, C. T., LOW, H. M., KIN SUNG, K. W., RIGOUTSOS, I., LORING, J. & WEI, C. L. 2010. Dynamic changes in the human methylome during differentiation. *Genome research,* 20**,** 320-31.

LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics, 11, 204-220.

LE GUEZENNEC, X., VERMEULEN, M., BRINKMAN, A. B., HOEIJMAKERS, W. A., COHEN, A., LASONDER, E. & STUNNENBERG, H. G. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol, 26, 843-51.

LEE, B. & MULLER, M. T. 2009. SUMOylation enhances DNA methyltransferase 1 activity. *The Biochemical journal,* 421**,** 449-61.

LEE, J.-H. & SKALNIK, D. G. 2002. CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. *The Journal of biological chemistry,* 277**,** 42259-42267.

LEE, J. S., SMITH, E. & SHILATIFARD, A. 2010. The language of histone crosstalk. *Cell,* 142**,** 682-5.

LEE, K. C. & LEE KRAUS, W. 2001. Nuclear receptors, coactivators and chromatin: new approaches, new insights. *Trends in endocrinology and metabolism: TEM,* 12**,** 191-197.

LEES-MURDOCK, D. J. & WALSH, C. P. 2008. DNA methylation reprogramming in the germ line. *Advances in experimental medicine and biology,* 626**,** 1-15.

LEI, H., OH, S. P., OKANO, M., JUTTERMANN, R., GOSS, K. A., JAENISCH, R. & LI, E. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development,* 122**,** 3195-205.

LEUNG, V. Y. L., GAO, B., LEUNG, K. K. H., MELHADO, I. G., WYNN, S. L., AU, T. Y. K., DUNG, N. W. F., LAU, J. Y. B., MAK, A. C. Y., CHAN, D. & CHEAH, K. S. E. 2011. SOX9 Governs Differentiation Stage-Specific Gene Expression in Growth Plate Chondrocytes via Direct Concomitant Transactivation and Repression. *PLoS Genetics,* 7**,** e1002356.

LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-26.

LI, G., LEVITUS, M., BUSTAMANTE, C. & WIDOM, J. 2004. Rapid spontaneous accessibility of nucleosomal DNA. *Nature Structural & Molecular Biology,* 12**,** 46-53.

LI, M., CHEN, D., SHILOH, A., LUO, J., NIKOLAEV, A. Y., QIN, J. & GU, W. 2002. Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature,* 416**,** 648-653.

LI, X., ITO, M., ZHOU, F., YOUNGSON, N., ZUO, X., LEDER, P. & FERGUSON-SMITH, A. C. 2008. A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell,* 15**,** 547-557.

LI, Y. Q., ZHOU, P. Z., ZHENG, X. D., WALSH, C. P. & XU, G. L. 2007. Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic acids research,* 35**,** 390-400.

LIM, H. G., SUZUKI, K., COOPER, D. A. & KELLEHER, A. D. 2008. Promoter-targeted siRNAs induce gene silencing of simian immunodeficiency virus (SIV) infection in vitro. *Molecular therapy : the journal of the American Society of Gene Therapy,* 16**,** 565-70.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q.-M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-322.

LIU, F., ZHAO, X., PERNA, F., WANG, L., KOPPIKAR, P., ABDEL-WAHAB, O., HARR, M. W., LEVINE, R. L., XU, H., TEFFERI, A., DEBLASIO, A., HATLEN, M., MENENDEZ, S. & NIMER, S. D. 2011. JAK2V617F-Mediated Phosphorylation of PRMT5 Downregulates Its Methyltransferase Activity and Promotes Myeloproliferation. *Cancer cell,* 19**,** 283-294.

LOUGHERY, J. E., DUNNE, P. D., O'NEILL, K. M., MEEHAN, R. R., MCDAID, J. R. & WALSH, C. P. 2011. DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. *Human molecular genetics,* 20**,** 3241-55.

LUCCI-CORDISCO, E. & NERI, G. 2009. Silent beginning: early silencing of the MED1/MBD4 gene in colorectal tumorigenesis. *Cancer biology & therapy,* 8**,** 192-193.

LUO, Y., LIN, F. T. & LIN, W. C. 2004. ATM-mediated stabilization of hMutL DNA mismatch repair proteins augments p53 activation during DNA damage. *Molecular and cellular biology,* 24**,** 6430-44.

LUPO, A., CESARO, E., MONTANO, G., IZZO, P. & COSTANZO, P. 2011. ZNF224: Structure and role of a multifunctional KRAB-ZFP protein. *The international journal of biochemistry & cell biology,* 43**,** 470-473.

LUSSER, A. & KADONAGA, J. T. 2003. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays,* 25**,** 1192-200.

MACKAY, D. J. G., CALLAWAY, J. L. A., MARKS, S. M., WHITE, H. E., ACERINI, C. L., BOONEN, S. E., DAYANIKLI, P., FIRTH, H. V., GOODSHIP, J. A., HAEMERS, A. P., HAHNEMANN, J. M. D., KORDONOURI, O., MASOUD, A. F., OESTERGAARD, E., STORR, J., ELLARD, S., HATTERSLEY, A. T., ROBINSON, D. O. & TEMPLE, I. K. 2008. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nature Genetics,* 40**,** 949-951.

MACKAY, J. P. & CROSSLEY, M. 1998. Zinc fingers are sticking together. *Trends in Biochemical Sciences,* 23**,** 1-4.

MADAPURA, M. P., SUTHERLAND, H. G., ULE, J., GRIMES, G. R. & BICKMORE, W. A. 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *Plos Genetics (in press).*

MANNERVIK, M. 1999. Transcriptional Coregulators in Development. *Science (New York, NY),* 284**,** 606-609.

MARTIN CABALLERO, I., HANSEN, J., LEAFORD, D., POLLARD, S. & HENDRICH, B. D. 2009. The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PloS one,* 4**,** e4315.

MATSUDA, H., PAUL, B. D., CHOI, C. Y., HASEBE, T. & SHI, Y.-B. 2009. Novel functions of protein arginine methyltransferase 1 in thyroid hormone receptor-mediated transcription and in the regulation of metamorphic rate in Xenopus laevis. *Molecular and Cellular Biology,* 29**,** 745-757.

MATSUMURA, T., SUZUKI, T., AIZAWA, K., MUNEMASA, Y., MUTO, S., HORIKOSHI, M. & NAGAI, R. 2005. The deacetylase HDAC1 negatively regulates the cardiovascular transcription factor Krüppel-like factor 5 through direct interaction. *The Journal of biological chemistry,* 280**,** 12123-12129.

MATZKE, M., MATZKE, A. J. & KOOTER, J. M. 2001. RNA: guiding gene silencing. *Science,* 293**,** 1080-3.

MATZKE, M., KANNO, T., HUETTEL, B., DAXINGER, L. & MATZKE, A. J. M. 2007. Targets of RNA-directed DNA methylation. Current Opinion in Plant Biology, 10, 512-519.

MCINNES, I. B. & SCHETT, G. 2011. The pathogenesis of rheumatoid arthritis. *The New England journal of medicine,* 365**,** 2205-2219.

MEISSNER, A., MIKKELSEN, T. S., GU, H., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X., BERNSTEIN, B. E., NUSBAUM, C., JAFFE, D. B., GNIRKE, A., JAENISCH, R. & LANDER, E. S. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,* 454**,** 766-770.

MENG, H. X., HACKETT, J. A., NESTOR, C., DUNICAN, D. S., MADEJ, M., REDDINGTON, J. P., PENNINGS, S., HARRISON, D. J. & MEEHAN, R. R. 2011. Apoptosis and DNA Methylation. *Cancers,* 3**,** 1798-1820.

METIVIER, R., GALLAIS, R., TIFFOCHE, C., LE PERON, C., JURKOWSKA, R. Z., CARMOUCHE, R. P., IBBERSON, D., BARATH, P., DEMAY, F., REID, G.,

BENES, V., JELTSCH, A., GANNON, F. & SALBERT, G. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature,* 452**,** 45-50.

MÉTIVIER, R., PENOT, G., HÜBNER, M. R., REID, G., BRAND, H., KOS, M. & GANNON, F. 2003. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell,* 115**,** 751-763.

MEULMEESTER, E., MAURICE, M., BOUTELL, C., TEUNISSE, A., OVAA, H., ABRAHAM, T., DIRKS, R. & JOCHEMSEN, A. 2005. Loss of HAUSP-Mediated Deubiquitination Contributes to DNA Damage-Induced Destabilization of Hdmx and Hdm2. *Molecular Cell,* 18**,** 565-576.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* 448**,** 553-60.

MILLAR, C. B. 2002. *Functional analysis of the methyl-CpG-binding DNA glycosylase MBD4.*

MILLAR, C. B., GUY, J., SANSOM, O. J., SELFRIDGE, J., MACDOUGALL, E., HENDRICH, B., KEIGHTLEY, P. D., BISHOP, S. M., CLARKE, A. R. & BIRD, A. 2002. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science,* 297**,** 403-5.

MIQUEL, C., JACOB, S., GRANDJOUAN, S., AIME, A., VIGUIER, J., SABOURIN, J. C., SARASIN, A., DUVAL, A. & PRAZ, F. 2007. Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. *Oncogene,* 26**,** 5919-26.

MIYAMOTO, S., SUZUKI, T., MUTO, S., AIZAWA, K., KIMURA, A., MIZUNO, Y., NAGINO, T., IMAI, Y., ADACHI, N., HORIKOSHI, M. & NAGAI, R. 2003. Positive and negative regulation of the cardiovascular transcription factor KLF5 by p300 and the oncogenic regulator SET through interaction and acetylation on the DNA-binding domain. *Molecular and Cellular Biology,* 23**,** 8528-8541.

MOREY, L. & HELIN, K. 2010. Polycomb group protein-mediated repression of transcription. *Trends in biochemical sciences,* 35**,** 323-32.

MORGAN, H. D. 2005. Epigenetic reprogramming in mammals. *Human Molecular Genetics,* 14**,** R47-R58.

MORTUSEWICZ, O., SCHERMELLEH, L., WALTER, J., CARDOSO, M. C. & LEONHARDT, H. 2005. Recruitment of DNA methyltransferase I to DNA repair sites. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 8905-9.

MOWEN, K. A., SCHURTER, B. T., FATHMAN, J. W., DAVID, M. & GLIMCHER, L. H. 2004. Arginine methylation of NIP45 modulates cytokine gene expression in effector T lymphocytes. Mol Cell, 15, 559-71.

MÜLLER, P., CROFTS, J. D., NEWMAN, B. S., BRIDGEWATER, L. C., LIN, C.-Y., GUSTAFSSON, J.-Å. & STRÖM, A. 2009. SOX9 mediates the retinoic acid-induced HES-1 gene expression in human breast cancer cells. *Breast Cancer Research and Treatment,* 120**,** 317-326.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. & ERLICH, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol,* 51 Pt 1**,** 263-73.

MUNZEL, M., LERCHER, L., MULLER, M. & CARELL, T. 2010. Chemical discrimination between dC and 5MedC via their hydroxylamine adducts. *Nucleic acids research,* 38**,** e192.

MUTO, M. 2002. Targeted Disruption of Np95 Gene Renders Murine Embryonic Stem Cells Hypersensitive to DNA Damaging Agents and DNA Replication Blocks. *Journal of Biological Chemistry,* 277**,** 34549-34555.

NAN, X., MEEHAN, R. R. & BIRD, A. 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research,* 21**,** 4886-4892.

NAN, X., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-9.

NARLIKAR, G. J., FAN, H.-Y. & KINGSTON, R. E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell,* 108**,** 475-487.

NESTOR, C., RUZOV, A., MEEHAN, R. & DUNICAN, D. 2010. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques,* 48**,** 317-9.

NESTOR, C. E., OTTAVIANO, R., REDDINGTON, J., SPROUL, D., REINHARDT, D., DUNICAN, D., KATZ, E., DIXON, J. M., HARRISON, D. J. & MEEHAN, R. R. 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome research,* 22**,** 467-77.

NG, H. H., JEPPESEN, P. & BIRD, A. 2000. Active Repression of Methylated Genes by the Chromosomal Protein MBD1. *Molecular and Cellular Biology,* 20**,** 1394-1406.

NG, H. H., ZHANG, Y., HENDRICH, B., JOHNSON, C. A., TURNER, B. M., ERDJUMENT-BROMAGE, H., TEMPST, P., REINBERG, D. & BIRD, A. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genetics,* 23**,** 58-61.

OGHAMIAN, S., SODIR, N. M., BASHIR, M. U., SHEN, H., CULLINS, A. E., CARROLL, C. A., KUNDU, P., SHIBATA, D. & LAIRD, P. W. 2011. Reduction of pancreatic acinar cell tumor multiplicity in Dnmt1 hypomorphic mice. *Carcinogenesis,* 32**,** 829-35.

OHKI, I., SHIMOTAKE, N., FUJITA, N., JEE, J., IKEGAMI, T., NAKAO, M. & SHIRAKAWA, M. 2001. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell,* 105**,** 487-97.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

OKANO, M., XIE, S. & LI, E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics,* 19**,** 219-20.

OOI, S. K., WOLF, D., HARTUNG, O., AGARWAL, S., DALEY, G. Q., GOFF, S. P. & BESTOR, T. H. 2010. Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics & chromatin,* 3**,** 17.

PAL, S. & SIF, S. 2007. Interplay between chromatin remodelers and protein arginine methyltransferases. *Journal of cellular physiology,* 213**,** 306-315.

PALII, S. S., VAN EMBURGH, B. O., SANKPAL, U. T., BROWN, K. D. & ROBERTSON, K. D. 2008. DNA methylation inhibitor 5-Aza-2'-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. *Molecular and cellular biology,* 28**,** 752-71.

PANDYA, A. Y. 2004. Nuclear Localization of KLF4 Is Associated with an Aggressive Phenotype in Early-Stage Breast Cancer. *Clinical Cancer Research,* 10**,** 2709-2719.

PASSERON, T., VALENCIA, J. C., NAMIKI, T., VIEIRA, W. D., PASSERON, H., MIYAMURA, Y. & HEARING, V. J. 2009. Upregulation of SOX9 inhibits the growth of human and mouse melanomas and restores their sensitivity to retinoic acid. *Journal of Clinical Investigation*.

PATEL, K., DICKSON, J., DIN, S., MACLEOD, K., JODRELL, D. & RAMSAHOYE, B. 2010. Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated

DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic acids research,* 38**,** 4313-24.

PATEL, S. R., BHUMBRA, S. S., PAKNIKAR, R. S. & DRESSLER, G. R. 2011. Epigenetic Mechanisms of Groucho/Grg/TLE Mediated Transcriptional Repression. *Molecular Cell***,** 1-11.

PETERS, A. H., O'CARROLL, D., SCHERTHAN, H., MECHTLER, K., SAUER, S., SCHOFER, C., WEIPOLTSHAMMER, K., PAGANI, M., LACHNER, M., KOHLMAIER, A., OPRAVIL, S., DOYLE, M., SIBILIA, M. & JENUWEIN, T. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell,* 107**,** 323-37.

PETERSON, C. L. & LOGIE, C. 2000. Recruitment of chromatin remodeling machines. *Journal of Cellular Biochemistry,* 78**,** 179-185.

PETRONZELLI, F., RICCIO, A., MARKHAM, G. D., SEEHOLZER, S. H., GENUARDI, M., KARBOWSKI, M., YEUNG, A. T., MATSUMOTO, Y. & BELLACOSA, A. 2000. Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J Cell Physiol,* 185**,** 473-80.

PICHLER, G., WOLF, P., SCHMIDT, C. S., MEILINGER, D., SCHNEIDER, K., FRAUER, C., FELLINGER, K., ROTTACH, A. & LEONHARDT, H. 2011. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *Journal of cellular biochemistry,* 112**,** 2585-93.

PIERA-VELAZQUEZ, S., HAWKINS, D. F., WHITECAVAGE, M. K., COLTER, D. C., STOKES, D. G. & JIMENEZ, S. A. 2007. Regulation of the human SOX9 promoter by Sp1 and CREB. *Experimental Cell Research,* 313**,** 1069-1079.

POLLACK, B. P., KOTENKO, S. V., HE, W., IZOTOVA, L. S., BARNOSKI, B. L. & PESTKA, S. 1999. The human homologue of the yeast proteins Skb1 and Hsl7p interacts with Jak kinases and contains protein methyltransferase activity. J Biol Chem, 274, 31531-42.

POPP, C., DEAN, W., FENG, S., COKUS, S. J., ANDREWS, S., PELLEGRINI, M., JACOBSEN, S. E. & REIK, W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature,* 463**,** 1101-1105.

PRITCHETT, J., ATHWAL, V., ROBERTS, N., HANLEY, N. A. & HANLEY, K. P. 2011. Understanding the role of SOX9 in acquired diseases: lessons from development. *Trends in Molecular Medicine,* 17**,** 166-174.

PROKHORTCHOUK, A., HENDRICH, B., JORGENSEN, H., RUZOV, A., WILM, M., GEORGIEV, G., BIRD, A. & PROKHORTCHOUK, E. 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development,* 15**,** 1613-8.

QIN, W., LEONHARDT, H. & SPADA, F. 2011. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *Journal of Cellular Biochemistry,* 112**,** 439-444.

QIU, C., SAWADA, K., ZHANG, X. & CHENG, X. 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature structural biology,* 9**,** 217-24.

QUENNEVILLE, S., VERDE, G., CORSINOTTI, A., KAPOPOULOU, A., JAKOBSSON, J., OFFNER, S., BAGLIVO, I., PEDONE, P. V., GRIMALDI, G., RICCIO, A. & TRONO, D. 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Molecular Cell,* 44**,** 361-372.

RAI, K., HUGGINS, I. J., JAMES, S. R., KARPF, A. R., JONES, D. A. & CAIRNS, B. R. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell,* 135**,** 1201-1212.

RAI, K., SARKAR, S., BROADBENT, T. J., VOAS, M., GROSSMANN, K. F., NADAULD, L. D., DEHGHANIZADEH, S., HAGOS, F. T., LI, Y., TOTH, R. K., CHIDESTER, S., BAHR, T. M., JOHNSON, W. E., SKLOW, B., BURT, R., CAIRNS, B. R. & JONES, D. A. 2010. DNA demethylase activity maintains intestinal cells in an undifferentiated state following loss of APC. *Cell,* 142**,** 930-942.

RAJAKUMARA, E., LAW, J. A., SIMANSHU, D. K., VOIGT, P., JOHNSON, L. M., REINBERG, D., PATEL, D. J. & JACOBSEN, S. E. 2011. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes & Development,* 25**,** 137-152.

RAMOCKI, N. M., WILKINS, H. R., MAGNESS, S. T., SIMMONS, J. G., SCULL, B. P., LEE, G. H., MCNAUGHTON, K. K. & LUND, P. K. 2007. Insulin Receptor Substrate-1 Deficiency Promotes Apoptosis in the Putative Intestinal Crypt Stem Cell Region, Limits Apcmin/+ Tumors, and Regulates Sox9. *Endocrinology,* 149**,** 261-267.

RAMSAHOYE, B. H., BINISZKIEWICZ, D., LYKO, F., CLARK, V., BIRD, A. P. & JAENISCH, R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 5237-42.

RAO, P., FULLER, G. N. & PRIETO, V. G. 2010. Expression of Sox-9 in metastatic melanoma--a potential diagnostic pitfall. *The American Journal of dermatopathology,* 32**,** 262-266.

REDDINGTON, J. 2011. A role for the DNA methylation system in polycomb protein-mediated gene regulation. *PhD thesis, The University of Edinburgh*.

REIK, W. 2001. Epigenetic Reprogramming in Mammalian Development. *Science (New York, NY),* 293**,** 1089-1093.

RICCIO, A., AALTONEN, L. A., GODWIN, A. K., LOUKOLA, A., PERCESEPE, A., SALOVAARA, R., MASCIULLO, V., GENUARDI, M., PARAVATOU-PETSOTAS, M., BASSI, D. E., RUGGERI, B. A., KLEIN-SZANTO, A. J., TESTA, J. R., NERI, G. & BELLACOSA, A. 1999. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nature genetics,* 23**,** 266-8.

RICLET, R., CHENDEB, M., VONESCH, J. L., KOCZAN, D., THIESEN, H. J., LOSSON, R. & CAMMAS, F. 2009. Disruption of the interaction between transcriptional intermediary factor 1{beta} and heterochromatin protein 1 leads to a switch from DNA hyper- to hypomethylation and H3K9 to H3K27 trimethylation on the MEST promoter correlating with gene reactivation. Mol Biol Cell, 20, 296-305.

RIGGS, A. D. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics,* 14**,** 9-25.

RIGGS, A. D. & JONES, P. A. 1983. 5-methylcytosine, gene regulation, and cancer. *Advances in cancer research,* 40**,** 1-30.

ROCHA, S. T. D., EDWARDS, C. A., ITO, M., OGATA, T. & FERGUSON-SMITH, A. C. 2008. Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in Genetics,* 24**,** 306-316.

ROWLAND, B. D. & PEEPER, D. S. 2005. KLF4, p21 and context-dependent opposing forces in cancer. *Nature Reviews Cancer,* 6**,** 11-23.

RUBIN, R. & STRAYER, D. 2012. Rubin's Pathology: Clinicopathologic Foundations of Medicine (Sixth Edition). *Wolters Kluwer*.

RUGO, R. E., MUTAMBA, J. T., MOHAN, K. N., YEE, T., CHAILLET, J. R., GREENBERGER, J. S. & ENGELWARD, B. P. 2011. Methyltransferases mediate cell memory of a genotoxic insult. *Oncogene,* 30**,** 751-6.

RUSSO, E., MARTIENSSEN, R. & RIGGS, A. D. 1996. Epigenetic Mechanisms of Gene Regulation. *Cold Spring Harbor Lab. Press, Plainview, NY*.

RUZOV, A., DUNICAN, D. S., PROKHORTCHOUK, A., PENNINGS, S., STANCHEVA, I., PROKHORTCHOUK, E. & MEEHAN, R. R. 2004. Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development,* 131**,** 6185-94.

RUZOV, A., SHORNING, B., MORTUSEWICZ, O., DUNICAN, D. S., LEONHARDT, H. & MEEHAN, R. R. 2009. MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development,* 136**,** 2277-2286.

SAITO, Y., KANAI, Y., SAKAMOTO, M., SAITO, H., ISHII, H. & HIROHASHI, S. 2001. Expression of mRNA for DNA methyltransferases and methyl-CpG-binding proteins and DNA methylation status on CpG islands and pericentromeric satellite regions during human hepatocarcinogenesis. *Hepatology,* 33**,** 561-8.

SAMPATH, S. C., MARAZZI, I., YAP, K. L., KRUTCHINSKY, A. N., MECKLENBRAUKER, I., VIALE, A., RUDENSKY, E., ZHOU, M. M., CHAIT, B. T. & TARAKHOVSKY, A. 2007. Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Molecular cell,* 27**,** 596-608.

SANSOM, O. J., ZABKIEWICZ, J., BISHOP, S. M., GUY, J., BIRD, A. & CLARKE, A. R. 2003. MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine. *Oncogene,* 22**,** 7130-6.

SARKARI, F., SANCHEZ-ALCARAZ, T., WANG, S., HOLOWATY, M. N., SHENG, Y. & FRAPPIER, L. 2009. EBNA1-mediated recruitment of a histone H2B deubiquitylating complex to the Epstein-Barr virus latent origin of DNA replication. *PLoS Pathog,* 5**,** e1000624.

SASAI, N. & DEFOSSEZ, P. A. 2009. Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *The International journal of developmental biology,* 53**,** 323-34.

SASAKI, H. & MATSUI, Y. 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Reviews Genetics,* 2008**,** 129-140.

SCARSDALE, J. N., WEBB, H. D., GINDER, G. D. & WILLIAMS, D. C., JR. 2011. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic acids research,* 39**,** 6741-52.

SCHLEGEL, J., GUNEYSU, S. & MENNEL, H. D. 2002. Expression of the genes of methyl-binding domain proteins in human gliomas. *Oncology reports,* 9**,** 393-5.

SCHLESINGER, Y., STRAUSSMAN, R., KESHET, I., FARKASH, S., HECHT, M., ZIMMERMAN, J., EDEN, E., YAKHINI, Z., BEN-SHUSHAN, E., REUBINOFF, B. E., BERGMAN, Y., SIMON, I. & CEDAR, H. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature genetics,* 39**,** 232-6.

SCHMITZ, K. M., MAYER, C., POSTEPSKA, A. & GRUMMT, I. 2010. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes & development,* 24**,** 2264-9.

SCHULTZ, D. C. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development,* 16**,** 919-932.

SCHULTZ, D. C., FRIEDMAN, J. R. & RAUSCHER, F. J. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Development,* 15**,** 428-443.

SCREATON, R. A., KIESSLING, S., SANSOM, O. J., MILLAR, C. B., MADDISON, K., BIRD, A., CLARKE, A. R. & FRISCH, S. M. 2003. Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link

between genome surveillance and apoptosis. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 5211-6.

SHALGI, R., PILPEL, Y. & OREN, M. 2010. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends in genetics : TIG,* 26**,** 253-9.

SHARIF, J., MUTO, M., TAKEBAYASHI, S., SUETAKE, I., IWAMATSU, A., ENDO, T. A., SHINGA, J., MIZUTANI-KOSEKI, Y., TOYODA, T., OKAMURA, K., TAJIMA, S., MITSUYA, K., OKANO, M. & KOSEKI, H. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature,* 450**,** 908-12.

SHARMA, S., KELLY, T. K. & JONES, P. A. 2010. Epigenetics in cancer. *Carcinogenesis,* 31**,** 27-36.

SHEARSTONE, J. R., POP, R., BOCK, C., BOYLE, P., MEISSNER, A. & SOCOLOVSKY, M. 2011. Global DNA Demethylation During Mouse Erythropoiesis in Vivo. *Science (New York, NY),* 334**,** 799-802.

SHILATIFARD, A. 2006. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual review of biochemistry,* 75**,** 243-69.

SILVA, F. P., HAMAMOTO, R., FURUKAWA, Y. & NAKAMURA, Y. 2006. TIPUH1 encodes a novel KRAB zinc-finger protein highly expressed in human hepatocellular carcinomas. *Oncogene,* 25**,** 5063-5070.

SMALLWOOD, A., BLACK, J. C., TANESE, N., PRADHAN, S. & CAREY, M. 2008. HP1-mediated silencing targets Pol II coactivator complexes. *Nature structural & molecular biology,* 15**,** 318-20.

SMALLWOOD, A., ESTEVE, P. O., PRADHAN, S. & CAREY, M. 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & development,* 21**,** 1169-78.

SONG, C. X., SZULWACH, K. E., FU, Y., DAI, Q., YI, C., LI, X., LI, Y., CHEN, C. H., ZHANG, W., JIAN, X., WANG, J., ZHANG, L., LOONEY, T. J., ZHANG, B., GODLEY, L. A., HICKS, L. M., LAHN, B. T., JIN, P. & HE, C. 2011a. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology,* 29**,** 68-72.

SONG, J., RECHKOBLIT, O., BESTOR, T. H. & PATEL, D. J. 2011b. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science,* 331**,** 1036-40.

SONG, J., TEPLOVA, M., ISHIBE-MURAKAMI, S. & PATEL, D. J. 2012. Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation. *Science (New York, NY),* 335**,** 709-712.

SONG, M. S., SALMENA, L., CARRACEDO, A., EGIA, A., LO-COCO, F., TERUYA-FELDSTEIN, J. & PANDOLFI, P. P. 2008. The deubiquitinylation and localization of PTEN are regulated by a HAUSP–PML network. *Nature,* 455**,** 813-817.

SPROUL, D., NESTOR, C., CULLEY, J., DICKSON, J. H., DIXON, J. M., HARRISON, D. J., MEEHAN, R. R., SIMS, A. H. & RAMSAHOYE, B. H. 2011. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4364-9.

SRIPATHY, S. P., STEVENS, J. & SCHULTZ, D. C. 2006. The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. *Molecular and Cellular Biology,* 26**,** 8623-8638.

STANCHEVA, I., HENSEY, C. & MEEHAN, R. R. 2001. Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *The EMBO journal,* 20**,** 1963-73.

STANCHEVA, I. & MEEHAN, R. R. 2000. Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes & development,* 14**,** 313-27.

STRAHL, B. D., BRIGGS, S. D., BRAME, C. J., CALDWELL, J. A., KOH, S. S., MA, H., COOK, R. G., SHABANOWITZ, J., HUNT, D. F., STALLCUP, M. R. & ALLIS, C. D. 2001. Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. Curr Biol, 11, 996-1000.

SUZUKI, K., SHIJUUKU, T., FUKAMACHI, T., ZAUNDERS, J., GUILLEMIN, G., COOPER, D. & KELLEHER, A. 2005. Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region. *Journal of RNAi and gene silencing : an international journal of RNA and gene targeting research,* 1**,** 66-78.

SUZUKI, T. 2005. Vascular Implications of the Kruppel-Like Family of Transcription Factors. *Arteriosclerosis, Thrombosis, and Vascular Biology,* 25**,** 1135-1141.

TACHIBANA, M., SUGIMOTO, K., NOZAKI, M., UEDA, J., OHTA, T., OHKI, M., FUKUDA, M., TAKEDA, N., NIIDA, H., KATO, H. & SHINKAI, Y. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & development,* 16**,** 1779-91.

TACHIBANA, M., UEDA, J., FUKUDA, M., TAKEDA, N., OHTA, T., IWANARI, H., SAKIHAMA, T., KODAMA, T., HAMAKUBO, T. & SHINKAI, Y. 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & development,* 19**,** 815-26.

TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.

TAIRA, K. 2006. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature,* 441**,** 1176.

TAN, C. P. & NAKIELNY, S. 2006. Control of the DNA methylation system component MBD2 by protein arginine methylation. Mol Cell Biol, 26, 7224-35.

TAN, J., YANG, X., ZHUANG, L., JIANG, X., CHEN, W., LEE, P. L., KARUTURI, R. K., TAN, P. B., LIU, E. T. & YU, Q. 2007. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes & development,* 21**,** 1050-63.

TANG, J., QU, L.-K., ZHANG, J., WANG, W., MICHAELSON, J. S., DEGENHARDT, Y. Y., EL-DEIRY, W. S. & YANG, X. 2006. Critical role for Daxx in regulating Mdm2. *Nature Cell Biology,* 8**,** 855-862.

TAO, Q., HUANG, H., GEIMAN, T. M., LIM, C. Y., FU, L., QIU, G. H. & ROBERTSON, K. D. 2002. Defective de novo methylation of viral and cellular DNA sequences in ICF syndrome cells. *Human molecular genetics,* 11**,** 2091-102.

TAYLOR, K. J., SIMS, A. H., LIANG, L., FARATIAN, D., MUIR, M., WALKER, G., KUSKE, B., DIXON, J. M., CAMERON, D. A., HARRISON, D. J. & LANGDON, S. P. 2010. Dynamic changes in gene expression in vivo predict prognosis of tamoxifen-treated patients with breast cancer. *Breast cancer research : BCR,* 12**,** R39.

TEE, W.-W., PARDO, M., THEUNISSEN, T. W., YU, L., CHOUDHARY, J. S., HAJKOVA, P. & SURANI, M. A. 2010. Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes & Development,* 24**,** 2772-2777.

TEFFERI, A., SKODA, R. & VARDIMAN, J. W. 2009. Myeloproliferative neoplasms: contemporary diagnosis using histology and genetics. *Nature reviews. Clinical oncology,* 6**,** 627-637.

THAPA, A., SHAHNAWAZ, M., KARKI, P., RAJ DAHAL, G., GOLAM SHAROAR, M., YUB SHIN, S., SUP LEE, J., CHO, B. & PARK, I.-S. 2008. Purification of inclusion body—forming peptides and proteins in soluble form by fusion to Escherichia coli thermostable proteins. *BioTechniques,* 44**,** 787-796.

TINI, M., BENECKE, A., UM, S.-J., TORCHIA, J., EVANS, R. M. & CHAMBON, P. 2002. Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Molecular Cell,* 9**,** 265-277.

TOYOTA, M. & SUZUKI, H. 2010. Epigenetic drivers of genetic alterations. *Advances in genetics,* 70**,** 309-23.

TSAI, H.-C. & BAYLIN, S. B. 2011. Cancer epigenetics: linking basic biology to clinical medicine. *Nature Publishing Group***,** 1-16.

TUDOR, M., AKBARIAN, S., CHEN, R. Z. & JAENISCH, R. 2002. Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 15536-41.

TURNER, B. M. 2007. Defining an epigenetic code. *Nature cell biology,* 9**,** 2-6.

UM, S., HARBERS, M., BENECKE, A., PIERRAT, B., LOSSON, R. & CHAMBON, P. 1998. Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *The Journal of biological chemistry,* 273**,** 20728-20736.

UNOKI, M., BRUNET, J. & MOUSLI, M. 2009a. Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochemical Pharmacology,* 78**,** 1279-1288.

UNOKI, M., KELLY, J. D., NEAL, D. E., PONDER, B. A., NAKAMURA, Y. & HAMAMOTO, R. 2009b. UHRF1 is a novel molecular marker for diagnosis and the prognosis of bladder cancer. *British journal of cancer,* 101**,** 98-105.

USHITA, M., SAITO, T., IKEDA, T., YANO, F., HIGASHIKAWA, A., OGATA, N., CHUNG, U., NAKAMURA, K. & KAWAGUCHI, H. 2009. Transcriptional induction of SOX9 by NF-κB family member RelA in chondrogenic cells. *Osteoarthritis and Cartilage,* 17**,** 1065-1075.

VAN DER HORST, A., DE VRIES-SMITS, A. M., BRENKMAN, A. B., VAN TRIEST, M. H., VAN DEN BROEK, N., COLLAND, F., MAURICE, M. M. & BURGERING, B. M. 2006. FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. *Nature cell biology,* 8**,** 1064-73.

VAN DER KNAAP, J. A., KUMAR, B. R., MOSHKIN, Y. M., LANGENBERG, K., KRIJGSVELD, J., HECK, A. J., KARCH, F. & VERRIJZER, C. P. 2005. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. *Molecular cell,* 17**,** 695-707.

VELASCO, G., HUBE, F., ROLLIN, J., NEUILLET, D., PHILIPPE, C., BOUZINBA-SEGARD, H., GALVANI, A., VIEGAS-PEQUIGNOT, E. & FRANCASTEL, C. 2010. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 9281-6.

WADDINGTON, C. H. 1942. 'The epigenotype'. *Endeavour***,** 18-20.

WADE, P. A., GEGONNE, A., JONES, P. L., BALLESTAR, E., AUBRY, F. & WOLFFE, A. P. 1999. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nature genetics,* 23**,** 62-6.

WAKEFIELD, R. I., SMITH, B. O., NAN, X., FREE, A., SOTERIOU, A., UHRIN, D., BIRD, A. P. & BARLOW, P. N. 1999. The solution structure of the domain from MeCP2 that binds to methylated DNA. *Journal of Molecular Biology,* 291**,** 1055-1065.

WANG, C., IVANOV, A., CHEN, L., FREDERICKS, W. J., SETO, E., RAUSCHER, F. J. & CHEN, J. 2005. MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation. *The EMBO Journal,* 24**,** 3279-3290.

WANG, H., LEAV, I., IBARAGI, S., WEGNER, M., HU, G. F., LU, M. L., BALK, S. P. & YUAN, X. 2008. SOX9 Is Expressed in Human Fetal Prostate Epithelium and Enhances Prostate Cancer Invasion. *Cancer Research,* 68**,** 1625-1630.

WANG, H., MCKNIGHT, N. C., ZHANG, T., LU, M. L., BALK, S. P. & YUAN, X. 2007. SOX9 Is Expressed in Normal Prostate Basal Cells and Regulates Androgen Receptor Expression in Prostate Cancer Cells. *Cancer Research,* 67**,** 528-536.

WASSENEGGER, M., HEIMES, S., RIEDEL, L. & SÄNGER, H. L. 1994. RNA-directed de novo methylation of genomic sequences in plants. Cell, 76, 567-576.

WEBER, M., HELLMANN, I., STADLER, M. B., RAMOS, L., PAABO, S., REBHAN, M. & SCHUBELER, D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics,* 39**,** 457-66.

WHITELAW, N. C., CHONG, S., MORGAN, D. K., NESTOR, C., BRUXNER, T. J., ASHE, A., LAMBLEY, E., MEEHAN, R. & WHITELAW, E. 2010. Reduced levels of two modifiers of epigenetic gene silencing, Dnmt3a and Trim28, cause increased phenotypic noise. *Genome biology,* 11**,** R111.

WILSON, B. G., WANG, X., SHEN, X., MCKENNA, E. S., LEMIEUX, M. E., CHO, Y. J., KOELLHOFFER, E. C., POMEROY, S. L., ORKIN, S. H. & ROBERTS, C. W. 2010. Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer cell,* 18**,** 316-28.

WINKELMANN, J., LIN, L., SCHORMAIR, B., KORNUM, B. R., FARACO, J., PLAZZI, G., MELBERG, A., CORNELIO, F., URBAN, A. E., PIZZA, F., POLI, F., GRUBERT, F., WIELAND, T., GRAF, E., HALLMAYER, J., STROM, T. M. & MIGNOT, E. 2012. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Human molecular genetics*.

WIZNEROWICZ, M., JAKOBSSON, J., SZULC, J., LIAO, S., QUAZZOLA, A., BEERMANN, F., AEBISCHER, P. & TRONO, D. 2007. The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. *J Biol Chem,* 282**,** 34535-41.

WONG, E., YANG, K., KURAGUCHI, M., WERLING, U., AVDIEVICH, E., FAN, K., FAZZARI, M., JIN, B., BROWN, A. M., LIPKIN, M. & EDELMANN, W. 2002. Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A,* 99**,** 14937-42.

WU, H. & ZHANG, Y. 2011. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev,* 25**,** 2436-52.

WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nature reviews. Molecular cell biology,* 11**,** 607-20.

WU, Z. L., ZHENG, S. S., LI, Z. M., QIAO, Y. Y., AAU, M. Y. & YU, Q. 2010. Polycomb protein EZH2 regulates E2F1-dependent apoptosis through epigenetically modulating Bim expression. *Cell death and differentiation,* 17**,** 801-10.

XI, Y., FORMENTINI, A., NAKAJIMA, G., KORNMANN, M. & JU, J. 2008. Validation of biomarkers associated with 5-fluorouracil and thymidylate synthase in colorectal cancer. *Oncology reports,* 19**,** 257-62.

XU, G. L., BESTOR, T. H., BOURC'HIS, D., HSIEH, C. L., TOMMERUP, N., BUGGE, M., HULTEN, M., QU, X., RUSSO, J. J. & VIEGAS-PEQUIGNOT, E. 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

YATES, J. R., RUSE, C. I. & NAKORCHEVSKY, A. 2009. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng,* 11**,** 49-79.

YILDIRIM, O., LI, R., HUNG, J.-H., CHEN, P. B., DONG, X., EE, L.-S., WENG, Z., RANDO, O. J. & FAZZIO, T. G. 2011. Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell,* 147**,** 1498-1510.

YOKOE, T., TOIYAMA, Y., OKUGAWA, Y., TANAKA, K., OHI, M., INOUE, Y., MOHRI, Y., MIKI, C. & KUSUNOKI, M. 2009. KAP1 Is Associated With Peritoneal Carcinomatosis in Gastric Cancer. *Annals of Surgical Oncology,* 17**,** 821-828.

YU, F., THIESEN, J. & STRÄTLING, W. H. 2000. Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. *Nucleic Acids Research,* 28**,** 2201-2206.

YU, X., TENG, H., MARQUES, A., ASHGARI, F. & IBRAHIM, S. M. 2009. High Resolution Mapping of Cia3: A Common Arthritis Quantitative Trait Loci in Different Species. *The Journal of Immunology,* 182**,** 3016-3023.

ZHANG, Y., NG, H. H., ERDJUMENT-BROMAGE, H., TEMPST, P., BIRD, A. & REINBERG, D. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes & Development,* 13**,** 1924-1935.

ZHANG, Y., ROHDE, C., TIERLING, S., JURKOWSKI, T. P., BOCK, C., SANTACRUZ, D., RAGOZIN, S., REINHARDT, R., GROTH, M., WALTER, J. & JELTSCH, A. 2009. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS genetics,* 5**,** e1000438.

ZHANG, Z. & TENG, C. T. 2003. Phosphorylation of Kruppel-like factor 5 (KLF5/IKLF) at the CBP interaction region enhances its transactivation function. *Nucleic Acids Research,* 31**,** 2196-2208.

ZHANG, H. & ZHU, J.-K. 2011. RNA-directed DNA methylation. Current Opinion in Plant Biology, 14, 142-147.

ZHAO, Q., RANK, G., TAN, Y. T., LI, H., MORITZ, R. L., SIMPSON, R. J., CERRUTI, L., CURTIS, D. J., PATEL, D. J., ALLIS, C. D., CUNNINGHAM, J. M. & JANE, S. M. 2009. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature Structural & Molecular Biology,* 16**,** 304-311.

ZUO, X., SHENG, J., LAU, H.-T., MCDONALD, C. M., ANDRADE, M., CULLEN, D. E., BELL, F. T., IACOVINO, M., KYBA, M., XU, G. & LI, X. 2012. Zinc Finger Protein ZFP57 Requires Its Co-factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain. *Journal of Biological Chemistry,* 287**,** 2107-2118.